6-30-2016

Structural And Functional Studies Of Nicotinamide Adenine Dinucleotide And Streptolysin S Biosynthesis Proteins From Streptococcus Pyogenes

William T. Booth
University of South Carolina

STRUCTURAL AND FUNCTIONAL STUDIES OF NICOTINAMIDE ADENINE DINUCLEOTIDE AND STREPTOLYSIN S BIOSYNTHESIS PROTEINS FROM *STREPTOCOCCUS PYOGENES*

BY

William T. Booth

Bachelor of Science
University of North Carolina at Greensboro, 2006

Master of Science
University of North Carolina at Greensboro, 2011

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Chemistry

College of Arts and Sciences

University of South Carolina

2016

Accepted by:

Maksymilian Chruszcz, Major Professor

James M. Sodetz, Committee Member

Michael D. Wyatt, Committee Member

Linda S.Shimizu, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

DEDICATION

Thank you to my immediate and extended family for your unconditional love, support, and prayers.

ACKNOWLEDGEMENTS

I would like to thank Dr. Maksymilian Chruszcz for sharing your wisdom with me. I admire your patience, as well as your passion for crystallography, and I am eternally grateful for the opportunity to work with you. I would like to thank Dr. Lukasz Lebioda for allowing me to work in his lab and introducing me to Dr. Leslie Lovelace and Dr. Lesa Offermann. Thank you both for your guidance, your patience, and showing me how to prepare my first crystal screen. Thank you to my fellow and former labmates: Dr. Nicholas Mank, Swanandi Pote, Caleb Schlachter, and Nikita Ussin for your support. I would like to acknowledge Dr. Thomas Makris, Steven Ratigan, Jose Amaya, Job Grant, and Chun Hsieh for their insight in preparation of cloning and kinetic studies. I would like to acknowledge my Undergraduate Assistants: David Mysona, Trevor Morris, Kathryn Clary, Taylor Karlin, Milan Shah, and Linda Taylor for all of you help, and I wish you all the best of luck in the future. Thank you to the Sloan Foundation for your financial support throughout my graduate studies. Thank you to Dr. Wayne Carver, Mrs. Ansley Roberts, and Mrs. Judy Lawrence for your support during my entrance into graduate studies through the Integrated Biomedical Science Program. Finally, I would like to thank my committee members: Dr. James Sodetz, Dr. Linda Shimizu, and Dr. Mike Wyatt for you guidance and approval toward attaining this degree.

Results shown in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center (19 BM/ ID) at the Advanced Photon Source. Argonne is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357

Use of LS-CAT (21 BM/ID) is supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P100081).

Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) (or 22 BM) beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html

Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357 and W-31-109-Eng-38.

ABSTRACT

Invasive infections caused by *Streptococcus pyogenes*, also known as Group A Strep (GAS), results in approximately 600,000 deaths annually. With evidence of antibiotic-resistant strains of this bacterium on the rise, there is a need for the identification of new drug targets to control these infections. In our approach we target the quinolinate-salvage pathway (QSP) and the streptolysin S (SLS) biosynthesis pathway. The QSP provides a secondary pathway for NAD^+ biosynthesis within this organism; the SLS pathway leads to the formation of a quorum sensing molecule (SLS). We hypothesize that inhibition of the pathways will lead to GAS cell death or will impair the growth of the bacterium.

This dissertation presents results of the functional and structural (using x-ray crystallography) characterization of select proteins involved in these pathways. Highlights of the information contained within each chapter are as follows:

- Chapter 1 provides background information on the QSP and SLS pathways and reports current research on both targets.
- Chapter 2 describes results from structural studies on quinolinate phosphoribosyltransferase (spNadC), of the QSP, and shows results from functional and structural studies, as well as the first structures of this protein and its deletion mutant ($spNadC_{\Delta 69A}$).
- Chapter 3 reports the results from functional characterization of nicotinate mononucleotide adenyltransferase (spNadD) of the QSP. Preliminary

structural studies are described. While the protein structure was not yet determined, homology modeling was used to provide insights into the proposed spNadD structure.

- Chapter 4 details the functional and structural studies on NAD^+ synthetase (spNadE) of the QSP. This protein, similarly to NadD, is considered to be a good drug target. This chapter describes the first structures of this protein determined in GAS. Structural results provide new details for the reaction mechanism and the conversion of the substrate (nicotinate adenine dinucleotide (NaAD)) into nicotinamide adenine dinucleotide (NAD^+) through channels identified at the dimer interface of spNadE.
- Chapter 5 describes results from molecular cloning, isolation, and refolding experiments conducted on select SLS biosynthesis proteins: SagB, SagC, SagD, and SagG. Structures of these proteins have not been determined; however, molecular cloning experiments for the co-expression of the SagBCD complex have shown promising results.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

ATP ..... Adenosine Triphosphate

APS ..... Advanced Photon Source

BLAST ..... Basic Local Alignment Search Tool

CDC ..... Centers for Disease Control and Prevention

DNA ..... Deoxyribonucleic Acid

DSF ..... Differential Scanning Fluorimetry

GAS ..... Group A Streptococcus

IMAC ..... Immobilized Metal Affinity Chromatography

IFN-γ ..... Interferon Gamma

LS-CAT ..... Life Sciences Collaborative Access Team

MALDI ..... Matrix Assisted Laser Desorption/Ionization

MBP ..... Maltose-Binding Protein

Mac1 ..... Immunoglobulin G degrading enzyme

NaAD ..... Nicotinate Adenine Dinucleotide

NAD^+ ..... Nicotinamide Adenine Dinucleotide

NaMN ..... Nicotinate Mononucleotide

PDB ..... Protien Data Bank

PHT ..... Pthalic Acid

PRPP ..... Phosphoribosyl Pyrophosphate

PPC ..... 5-Phosphoribosyl-1-(Beta-Methylene) Pyrophosphate

QA ..... Quinolinic Acid

QSE ..... Quinolinate Salvage Enzyme

QSP ..... Quinolinate-Salvage Pathway

sag operon ..... Stretolysin-Associated Gene Operon

SagA ..... Streptolysin precurson

SagB ..... Proposed Dehydrogenase

SagBCD ..... Proposed Streptolysin Biosynthesis Complex

SagC ..... Proposed Cyclogenase

SagD ..... Proposed Dehydrogenase/ Cyclogenase Linker Protein

SagE ..... Proposed Group A Strep Immunity Protein

SagG ..... ATP Binding Cassete Protein 1

SagH ..... ATP Binding Cassete Protein 2

SagI ..... ATP Binding Cassete Protein 3

SBC-CAT ..... Structural Biology Center Collaborative Access Team

ScpA ..... Complement 5A peptidase

SER-CAT ..... Southeastern Regional Collaborative Access Team

Sla ..... Multidrug Resistance

SLS ..... Streptolysin S

SLO ..... Streptolysin O

SpNadC ..... Phosphoribosyltransferase from *Streptococcus pyogenes*

SpNadD ..... Nicotinate Mononucleotide Transferase from *Streptococcus pyogenes*

SpNadE ..... NAD^+ Synthetase from *Streptococcus pyogenes*

SpyCEP ..... Streptococcal Chemokine Protease

CHAPTER 1

INTRODUCTION TO *STREPTOCOCCUS PYOGENES* AND THE BIOSYNTHESIS OF NICOTINAMIDE ADENINE DINUCLEOTIDE AND STEPTOLYSIN S

1.1 BACKGROUND AND INTRODUCTION TO *STREPTOCOCCUS PYOGENES*

Streptococcus pyogenes, also known as Group A Streptococcus (GAS), is a virulent, β-hemolytic, bacterium (1). It is a gram positive, anaerobic, spore-forming bacterium that requires nutrient rich environments for optimal growth. This fact makes the human body an ideal host for this organism (2). GAS is responsible for the onset of infections associated with acute pharyngitis (strep throat), rheumatic fever, scarlet fever, toxic shock syndrome, and necrotizing fasciitis (2, 3).

1.2 GLOBAL IMPACT

Globally, severe cases of GAS diseases cause over 500,000 deaths per year, which places GAS among the ten most common causes of death due to human pathogens (4). In less developed countries, specifically Papua New Guinea, Ethiopia, Gambia, and the Philippines, invasive GAS infections attribute to approximately 663,000 new cases a year. An estimated 160,000 deaths of children, 90 days and younger, make up one third of the fatalities (4). In children 15 years and younger, 25% of GAS infections are considered fatal and a higher percentage is assumed in undocumented, lesser developed communities (4).

The CDC reports that penicillin is still an effective antibiotic for treatment of GAS infection, but macrolide, tetracycline, and clindamycin-resistant strains have been identified (5). As a result of these findings, new antibiotics must be created to treat these resistant strains. This dissertation will explain two new strategies for antibiotic development through the use of macromolecular structure-based drug design.

1.3 The Quinolinate-Salvage Pathway and NAD^+ Biosynthesis

Previous studies have identified that non-specific exotoxin release from GAS during infection promotes the activation of interferon-gamma (IFN-γ) (6). IFN-γ aids in the initiation of tryptophan degradation by monocytes (blood)/macrophages (tissue) by way of the kynurenine pathway (6, 7). The end products of this pathway are quinolinic acid (QA) and picolinic acid (8). It could be presumed that there would be increased levels of tryptophan degradation and consequently, high levels of quinolinate at the site of infection due to immune response.

GAS has evolved to utilize quinolinate, by the action of the quinolinate-salvage pathway (QSP) (1). This ability affords these bacterial species the capacity to use this molecule for essential biochemical processes where similar streptococcal species could not. Moreover, this adaptation improves GAS's odds of survival and could promote an increased virulence when compared to other opportunistic bacteria. Previous studies have identified that this pathway is responsible for the utilization of quinolinic acid for the biosynthesis of nicotinamide adenine dinucleotide (NAD^+) (1). NAD^+ is a necessary cofactor in the progression of numerous biochemical pathways including glycolysis, gluconeogenesis, the citric acid cycle, and the electron transport chain (9). The QSP is composed of quinolinate-salvage enzymes (QSEs): quinolinate phosphoribosyltansferase

(spNadC), nicotinate mononucleotide transferase (spNadD), and NAD synthetase (spNadE), which are collectively responsible for the conversion of quinolinate (QA), nicotinate/ niacin (Na), and nicotinamide (Nm) into NAD^+ (Figure 1.1) (1). Specifically, for the utilization of QA, NadC is responsible for the conversion of phosphoribosyl pyrophosphate (PRPP) and QA into nicotinate mononucleotide (NaMN) (1). This transformation is regarded as the first of three biosynthetic steps in the formation of NAD^+ (1). The two remaining biosynthetic steps are completed by the action of NadD and NadE (1). NadD is responsible for the conversion of NaMN and Adenosine triphosphate (ATP) into nicotinic acid adenine dinucleotide (NaAD) and NadE is responsible for the transfer of ammonia to convert NaAD into NAD^+ (1, 10). Detailed explanations of QSE activity will be explained in their respective chapters. Figure 1.2 provides an overview of the reactions catalyzed by the activity of the QSEs.

1.4 Virulence and Streptolysin S Biosynthesis

GAS comes in a number of different forms known as serotypes (11). These serotypes (M1-6, M12, M18, M28, M49) are all descriptions of minor genetic variations between different strains of GAS (11). The M designation comes from a product of the bacteria's *emm* gene referred to as the "M-protein" (11). This protein is a major factor in the classification and determination of the virulence of GAS strains. Other factors include streptococcal phospholipase (Sla), which is responsible for promoting adherence of GAS to epithelial tissue, streptococcal chemokine protease (SpyCEP), which promotes the destruction of interleukin signaling proteins, and immunoglobulin G degrading enzyme (Mac1), which is considered to be responsible for antibody degradation (11, 12). Furthermore, additional components are streptococcal inhibitor of complement (Sic) and

complement 5A peptidase (ScpA), both of which inhibit the activity of the complement system, and streptolysin O and S (SLO and SLS) which utilize different mechanisms for the lysis of different cell types (11).

SLS can promote the lysis of red and white blood cells, platelets, and subcellular organelles (2, 11). As a result, GAS gains access to nutrients and evades the body's immune defenses (11). Supplemented by the activity of the virulence proteins described above, the immune system manifests the site of infection as a location that looks as if skin has been eaten away (11). This description triggered the "flesh-eating bacteria" moniker, which is commonly used to describe GAS-related necrotizing fasciitis.

1.5 Relevance and Hypotheses

Severe GAS infections, like necrotizing fasciitis, are serious disorders and if left untreated can result in gruesome and disfiguring maladies. Severe infections can lead to amputations and even death. The general focus of this research resides in understanding the means by which GAS promotes infection and contributes to onset of necrotizing fasciitis. We presume that the inhibition of SLS and NAD^+ biosynthesis can be accomplished in order to prevent the progression of infection.

More specifically, focus of this is work is two-fold:

1. We are interested in the isolation and structural determination of the proteins involved in the activity of the QSP for structure-based inhibitor design. We hypothesize that through the inhibition of the QSEs, we can stop bacterial growth.
2. We are interested in the isolation and structural determination of the proteins involved in the biosynthesis and release of Streptolysin S. We hypothesize that through the inhibition of proteins involved in the biosynthesis of SLS (SagB-D)

and its release (SagG) we can prevent SLS formation, and/or stop the progression of necrotizing fasciitis.

1.6 Current research into the QSP and NAD+ Biosynthesis

According to Sorci *et al.*, the QSP operates as a secondary pathway in the formation of NAD^+ (1). The capacity to express spNadC is the key factor in this process. Genomic studies of streptococcal species revealed that *nadC* is present only in *Streptococcus pyogenes* and *Streptococcus pneumoniae* and may be related to the increased virulence as compared to other, less invasive, streptococci (1).

In opportunistic streptococcal species, e.g. *Streptococcus mutants* and *Streptococcus agalactiae*, the process of NAD^+ biosynthesis starts at the introduction of Na and Nm by the action of the NiaX membrane transporter. NiaX has been previously identified as a Na transporter which allows for the transport of Na and Nm across the cell membrane into the cytoplasm (1). The next biosynthetic step is the conversion of Na into NaMN by the action of nicotinate phosphoribosyl transferase (PncB) (1). Prior to the conversion to NamN, Nm must be converted to Na before NaMN biosynthesis can occur (1). The conversion is completed by the action of PncA nicotidaminase (PncA) which substitutes a hydroxyl group for an amine group before conversion to NaMN by PncB (1). As described above, other opportunistic bacteria also encode for NadD which converts NaMN into NaAD, and NadE which converts NaAD into NAD^+ (1). Refer to Figure 1.8.1 for a general schematic of this process.

Avirulent streptococci, e.g. *Streptococcus equi* and *Streptococcus suis*, utilize a more simplified approach to NAD^+ biosynthesis (1). After the transport of nicotinamide

riboside (NmR) into the cell by PnuC (an NmR transporter), NadR biochemically modifies NmR by operating as a dual-action enzyme. Its first action is to operate as a nucleoside kinase converting NmR to nicotinamide mononucleotide (NMN); the second action is as an adenyltransferase converting NMN to NAD^+ (13). Figure 1.1 also provides a general description of this process.

What separates GAS and *S. pneumonia* from the other streptococcal species is their ability to utilize a third pathway in combination with the pathway used by opportunistic species, described above (1). Both species encode for a NadC and convert QA, which is produced during infection, into NaMN for use by NadD, and eventually NadE, for the formation of NAD^+ (1). Original speculation was that QA was produced internally by the bacteria through aspartate metabolism by L-aspartate oxidase (NadB), and iminoaspartate metabolism by quinolinic acid synthetase (NadA) was proposed to lead to formation of QA as an end-product (1). It was concluded from knockout studies of PncA and NadC that NadC is obtained from the extracellular environment (1). A simplified description of this process can also be found in Figure 1.1.

In knockout studies of NadD and NadE homologs in *Mycobacterium* species, it was discovered that NadD and NadE were essential for the survival of the species, and their loss diminished NAD^+ biosynthesis intermediate production and lead to cell death (14)

In this dissertation, the first published structures of NadC and NadE will be presented. To date, there is currently no structural data on any of the QSE's from *S. pyogenes*.

1.7 CURRENT RESEARCH INTO STREPTOLYSIN S (SLS) BIOSYNTHESIS

SLS is a bacteriocin, which is synthesized from the collective activity of the *streptolysin associated gene* operon (*sag* operon) of the *S. pyogenes* genome (2, 15). The *sag* operon is a collective of nine open reading frames (ORF's) referred to as *sagA* – *sagI* (15). All the ORFs encode for proteins necessary for the biosynthesis and export of SLS (2). The *sagA* gene encodes for a 53 amino acid strand (SagA) that acts as a precursor for the formation of SLS (2, 15); upon modification by operon proteins, it forms mature SLS (2, 16). SagB is proposed to act as a dehydrogenase working in conjunction with SagC and SagD, which allows for the withdrawal of necessary hydrogens in preparation for mature SLS formation (16). SagC is suggested to operate as a cyclodehydratase that plays a role in the conversion of cysteine, serine, and threonine residues of immature SLS into oxazole, thiazole, and methyloxazole, respectively, to form mature SLS (16). SagD is proposed to be responsible for the formation of the SagBCD complex and functions in the regulation of the complex's activity (16). SagE has been reported to have two roles: to act as a peptidase to promote the removal of an inactive peptide strand of SagA (approximately 23 amino acids), and to maintain an immunity role in GAS (2). SagG is proposed to act as a membrane-associated protein responsible for the binding of ATP (2). SagH and SagI work in conjunction with SagG, and are considered to be responsible for the release of mature SLS into its extracellular environment (2). SagF is believed to be a membrane-associated protein, but its function has yet to be defined (2). Figure 1.3 and Table 1.1 provides a schematic of the activity of the sag operon, as well as a description of the proposed function of each of the proteins involved.

At the inception of this project in 2012, the Protein Data Bank (PDB) contained no other homologous structures with significant sequence similarity, to SagB, SagC, SagD, or SagG, over the entire sequence (Table 1.2). This observation indicates the possibility of identifying a new enzyme, which could lead to new opportunities for future studies.

1.8 FIGURES AND TABLES

Table 1.1**:** The Sag proteins. This table shows proteins encoded by the *sag* operon, their molecular weights, and their proposed function (2).

Protein	# of Amino Acids	Molecular Weight (kDa)	Proposed Function
SagA	53	5	Amino acid precursor of Streptolysin S
SagB	316	36	Dehydrogenase
SagC	352	40	Cyclogenase
SagD	452	51	Linker of SagBC?
SagE	220	25	Immunity role/ Truncation of SagA?
SagF	227	26	Membrane associated protein?
SagG	307	41	ATP Binding Cassette transporter (ABC transporter)
SagH	375	42	ABC transporter
SagI	372	41	ABC transporter

Table 1.2: Sag proteins and their homologs in PDB. These are **r**esults of a sequence based search for proteins similar to SagB, SagC, SagD, and SagG with structures reported to Protein Data Bank. The results show very low sequence similarity of SagB-G to proteins with known structures (19).

Protein	# of amino acids	The most similar protein with known structure as identified with BLAST	Sequence identity (%)	Length of the similar protein fragment (# amino acids)
SagB	316	*Thermotoga maritima*, bacterioferritin comigratory protein PDB code: 4EO3	18	57
SagC	352	*Escherichia coli* MccB protein PDB code: 3H9G	10	35
SagD	452	*Prochloron sp. 06037A* heterocyclase TruD PDB code: 4BS9	23	103
SagG	307	*Thermotoga maritima* Abc Transporter Atp-binding Protein PDB code: 4YER	41	125

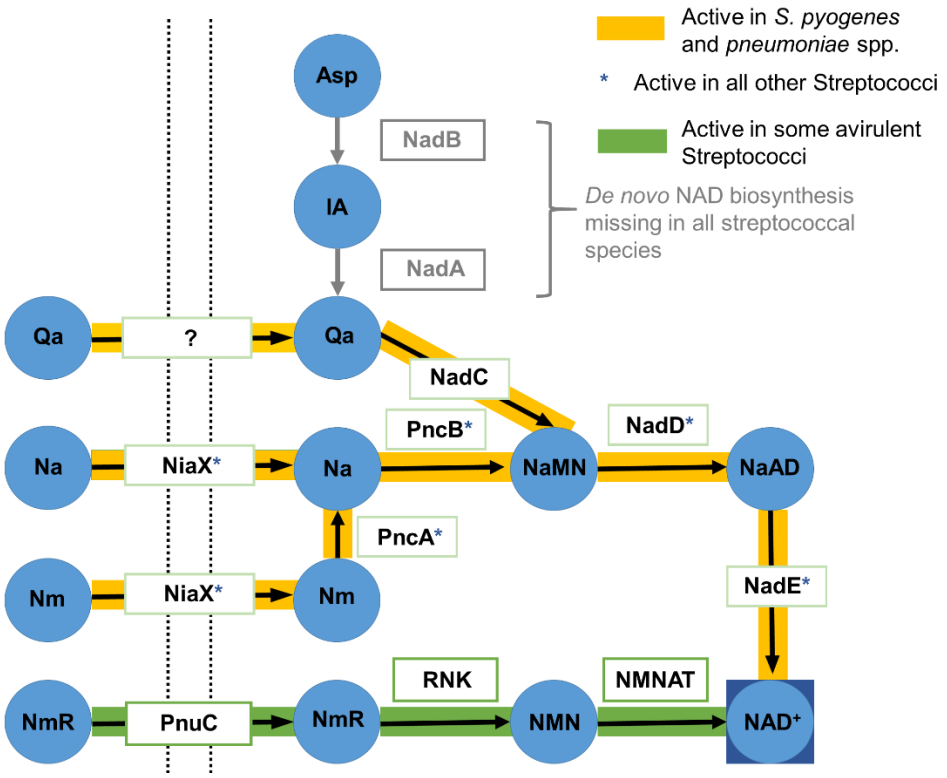


Figure 1.1: Integrated pathway for NAD^+ biosynthesis in streptococcal species. This schematic diagram, adapted from Sorci *et al.*, displays proteins and intermediates involved in NAD^+ biosynthesis in all Streptococcal species (1). Virulent species of Streptococci, specifically *S. pneumoniae* and *S. pyogenes*, exhibit the ability to utilize quinolinate as an alternate route for nicotinate mononucleotide (NaMN) production which may provide an explanation for the increased virulence of these species (1).

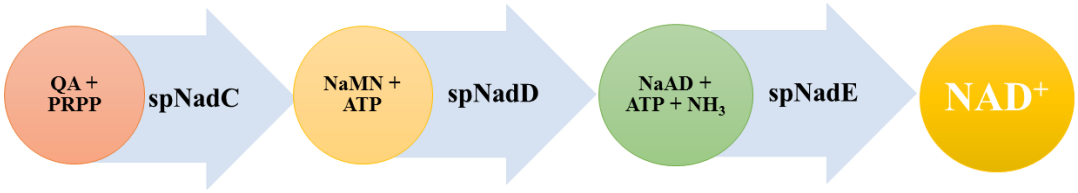
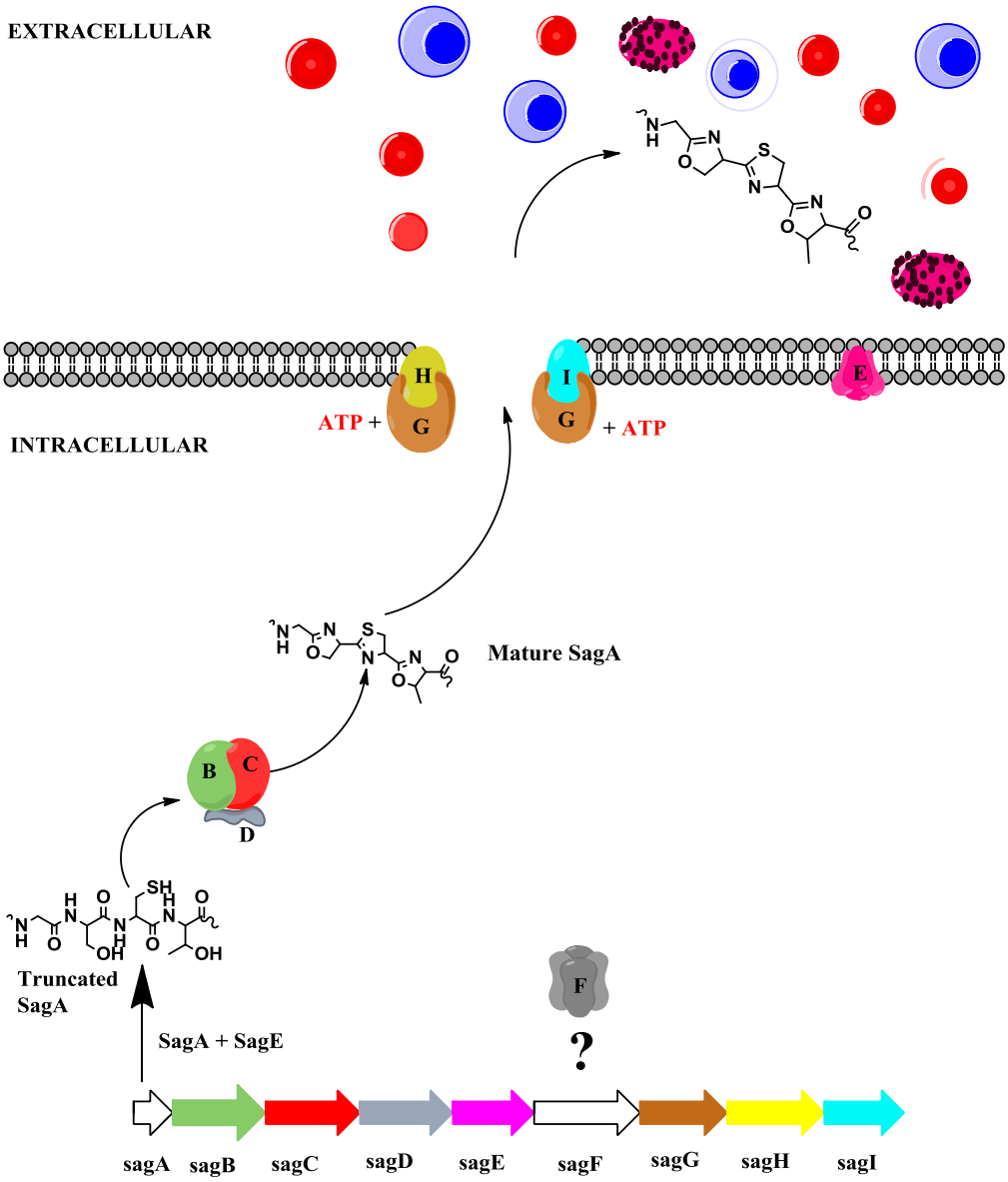


Figure 1.2: The quinolinate-salvage pathway. This figure displays a general description of the enzymatic activities of the proteins of the QSP (1).



THE SAG OPERON

Figure 1.3: The *sag* operon. This figure, adapted from Molloy *et al.*, provides an overview of the proposed activity of the *sag* operon during infection. SagA is converted into SLS by the collective activity of the the SagBCD complex and SagE (2, 16). SagGHI work together to form the proposed ATP binding cassette responsible for the transfer of SLS from the intracellular environment to the extracellular environment (2). Once SLS exits the cell it can cause the destruction of red blood cells, white blood cells, and platelets. SagF is labeled with a question mark because the function of the protein has yet to be defined. (2)

1.9 REFERENCES:

1. Sorci, L., I. K. Blaby, I. A. Rodionova, J. De Ingeniis, S. Tkachenko, V. De Crecy-Lagard, and A. L. Osterman. "Quinolinate Salvage and Insights for Targeting NAD Biosynthesis in Group A Streptococci." *Journal of Bacteriology* 195.4 (2012): 726-32. Web.
2. Molloy, Evelyn M., Paul D. Cotter, Colin Hill, Douglas A. Mitchell, and R. Paul Ross. "Streptolysin S-like Virulence Factors: The Continuing SagA." *Nature Reviews Microbiology Nat Rev Micro* 9.9 (2011): 670-81. Web.
3. Bauman, Robert W., Elizabeth Machunis-Masuoka, and Ian R. Tizard. Microbiology. San Francisco:Pearson/Benjamin Cummings, 2004.
4. Carapetis, Jonathan R., Andrew C. Steer, E. Kim Mulholland, and Martin Weber. "The Global Burden of Group A Streptococcal Diseases." *The Lancet Infectious Diseases* 5.11 (2005): 685-94. Web.
5. Rubio-López, Virginia, Sylvia Valdezate, David Álvarez, Pilar Villalón, María Medina, Celia Salcedo, and Juan-Antonio Sáez-Nieto. "Molecular Epidemiology, Antimicrobial Susceptibilities and Resistance Mechanisms of Streptococcus Pyogenes Isolates Resistant to Erythromycin and Tetracycline in Spain (1994–2006)."*BMC Microbiology BMC Microbiol* 12.1 (2012): 215. Web.
6. Murr, Christian, Dieter Gerlach, Bernhard Widner, Manfred P. Dierich, and Dietmar Fuchs. "Neopterin Production and Tryptophan Degradation in Humans Infected by Streptococcus Pyogenes." *Medical Microbiology and Immunology* 189.3 (2001): 161-63. Web.
7. Heyes, M. P., K. Saito, and S. P. Markey. "Human Macrophages Convert L-tryptophan into the Neurotoxin Quinolinic Acid." *Biochem. J. Biochemical Journal* 283.3 (1992): 633-35. Web.
8. Grant, R. S., S. E. Coggan, and G. A. Smythe. "The Physiological Action of Picolinic Acid in the Human Brain." *International Journal of Tryptophan Research* 2 (2009): 71-79. Web
9. Voet, Donald, and Judith G. Voet. *Biochemistry*. 4th ed. Hoboken, NJ: Wiley, 2011. Print. pgs 585-823
10. Cao, Hong, Beth L. Pietrak, and Charles Grubmeyer. "Quinolinate Phosphoribosyltransferase: Kinetic Mechanism for a Type II PRTase †." *Biochemistry* 41.10 (2002): 3520-528. Web.
11. Olsen, Randall J., and James M. Musser. "Molecular Pathogenesis of Necrotizing Fasciitis." *Annual Review of Pathology: Mechanisms of Disease* 5.1 (2010): 1-31. Print.
12. Zingaretti, C., F. Falugi, V. Nardi-Dei, G. Pietrocola, M. Mariani, S. Liberatori, M. Gallotta, M. Tontini, C. Tani, P. Speziale, G. Grandi, and I. Margarit. "Streptococcus Pyogenes SpyCEP: A Chemokine-inactivating Protease with Unique Structural and Biochemical Features." *The FASEB Journal* 24.8 (2010): 2839-848. Print
13. Grose, J. H., U. Bergthorsson, and J. R. Roth. "Regulation of NAD Synthesis by the Trifunctional NadR Protein of Salmonella Enterica." *Journal of Bacteriology* 187.8 (2005): 2774-782. Web.
14. Rodionova, I. A., B. M. Schuster, K. M. Guinn, L. Sorci, D. A. Scott, X. Li, I. Kheterpal, C. Shoen, M. Cynamon, C. Locher, E. J. Rubin, and A. L. Osterman.

"Metabolic and Bactericidal Effects of Targeted Suppression of NadD and NadE Enzymes in Mycobacteria." *MBio* 5.1 (2014): n. pag. Web.
15. Nizet, V., B. Beall, D. J. Bast, V. Datta, L. Kilburn, D. E. Low, and J. C. S. De Azavedo. "Genetic Locus for Streptolysin S Production by Group A Streptococcus." *Infection and Immunity* 68.7 (2000): 4245-254. Web.
16. Mitchell, Douglas A., Shaun W. Lee, Morgan A. Pence, Andrew L. Markley, Joyce D. Limm, and Victor Nizet. "Structural and Functional Dissection of the Heterocyclic Peptide Cytotoxin Streptolysin S." *Journal of Biolgical Chemistry* 284.119 (2009): 13004-3012. Print.
17. Cserzo, M., E. Wallin, I. Simon, G. Von Heijne, and A. Elofsson. "Prediction of Transmembrane Alpha-helices in Procariotic Membrane Proteins: The Dense Alignment Surface Method." *Protein Engineering* 10.6 (1997): 673-76. Print.
18. Hirokawa, T., S. Boon-Chieng, and S. Mitaku. "SOSUI: Classification and Secondary Structure Prediction System for Membrane Proteins." *Bioinformatics* 14.4 (1998): 378-79. Print.
19. Magrane, M., and Uniprot Consortium. "UniProt Knowledgebase: A Hub of Integrated Protein Data." *Database* 2011 (2011): 2-13. Print.

CHAPTER 2

STRUCTURAL AND FUNCTIONAL STUDIES OF QUINOLINATE PHOSPHORIBOSYLTRANSFERASE FROM *STREPTOCOCCUS PYOGENES*

2.1 INTRODUCTION

Quinolinate phosphoribosyltransferase (spNadC) is responsible for the first biosynthetic step in a triad of steps related to the formation of NAD^+, by way of the QSP, in GAS (1). This enzyme is specifically responsible for the N-nucleoside bond formation between PRPP and QA in order to form NaMN (2). This ability to use QA to biosynthesize NAD^+ enables GAS to use more than one biosynthetic route for NAD^+ production. Therefore, spNadC is also not considered a drug target, since GAS can also use NiaX membrane proteins in order to utilize nicotinate and nicotinamide for NaMN biosynthesis (1). The ability to utilize various pathways for NAD^+ biosynthesis most likely improves survival of GAS during infection (1). The purpose of this work is to understand the effect of spNadC and its role in promoting virulence in GAS.

The genomic studies presented by Sorci *et al.* indicated that GAS and *Streptococcus pneumoniae* (*S. pneumoniae*) are the only two streptococcal species that have the ability to produce NadC. It was also suggested that the presence of NadC may be related to the increased virulence of *S. pyogenes* and *S. pneumoniae* when compared to other avirulent and opportunistic streptococcal species (1). Several NadC homologs have their structures

determined. They have been identified in eukaryotic species including humans (hsNadC; PDB code: 3LAR; 40% sequence identity; RMSD: 1.4Å over 264 Cα atoms) and *Saccharomyces cerevisiae* (scNadC; PDB code: 3C2R; 35% sequence identity; RMSD: 1.7 Å over 247 Cα atoms). Pathogenic, non-streptococcal species for which a structure of spNadC homolog was determined include: *Vibrio cholerae* (vcNadC; PDB code: 3PAJ; 45% sequence identity; RMSD: 1.5 Å over 267 Cα atoms), *Salmonella enterica* (stNadC; PDB code: 1QAP, 41% sequence identity; RMSD: 1.9 Å over 262 Cα atoms), *Mycobacterium tuberculosis* (mtNadC; PDB code: 1QPR; 40% sequence identity; RMSD: 1.3 Å over 272 Cα atoms), *Helicobacter pylori* (hpNadC; PDB code: 2B7N; 38 % sequence identity; RMSD: 1.5 Å over 262 Cα atoms), and *Franciscella tularensis* (ftNadC; PDB code: 3TQV: 36 % sequence identity; RMSD 1.4 Å over 262 Cα atoms).

According to the Pfam database, NadC is a part of a family of proteins that have an αβ hydrolase fold (3, 4). This fold commonly contains a central parallel β sheet that is composed of several beta strands which is flanked by α-helices. NadC belongs to the TIM-phosphate binding superfamily of proteins that are grouped by the presence of a TIM-barrel structure containing the same 8 α helical/ 8 β strand feature described above in the αβ hydrolase family (5, 6). The barrel-type fold corresponds to the C-terminal part of NadC, while the N-terminal part of the protein contains an anti-parallel β sheet that is flanked on one side by four helices (Figure 2.1).

This chapter will present the first structural information on native spNadC and $spNadC_{\Delta 69A}$ deletion mutant.

2.2 Results and Discussion

Protein Production

All QSEs were over-expressed and purified in high quantities. SpNadC isolation resulted in approximately 240 mg of the protein per 1L of culture, as measured after an immobilized metal affinity chromatography (IMAC) step. Protein purity was confirmed using both SDS-PAGE and mass spectrometry. SpNadC contains a purification tag composed of 25 amino acids that encodes for a tobacco etch virus (TEV) cleavage site. Cleavage of this tag, with TEV however, was unsuccessful. SpNadC was further purified by gel filtration, and the elution profile contained one single peak that corresponds to a molecular weight of ~200 kDa. The same results were obtained for both the native and mutated protein. Dynamic light scattering was also used to determine the molecular weight of $NadC_{\Delta 69A}$ oligomer in solution. Results from these experiments indicate the molecular weight of NadC at 212 ± 80 kDa. These results indicate that most likely in solution spNadC forms a hexamer.

Kinetic Data

According to Sharma *et al.*, mtNadC is proposed to operate in an ordered-sequential mechanism with QA binding first and PRPP binding afterward (2) (Figure 2.2). It was assumed that spNadC operates according to the same mechanism.

Following a strategy adapted from Cao *et al*., an activity assay was conducted for native spNadC, as well as $spNadC_{\Delta 69A}$ (7). The assay did show activity for both constructs as evidenced by increasing absorbance values, at 266 nm, over a 10 minute span. Our results indicate that spNadC did not follow classic Michaelis-Menten kinetics, but did show

increased NaMN production with increasing PRPP concentrations from 0.2 mM – 1.6 mM. An inhibitory result was identified as the QA concentration was increased from 2-20 µM. It should be noted that the QA concentration is significantly smaller, in this experiment, compared to the PRPP experiments. The inhibitory outcome was also not observed with kinetic results from homologs in the literature (2, 7).

Initial spNadC kinetics studies indicate that this enzyme is very inefficient, which is consistent with previous studies (1). However, in the case of the protein used in this study, it was observed that increased concentration of QA leads to enzyme inhibition. The differences in the results of kinetics studies may be related to the presence of the purification tag in the protein that was produced. Recombinant NadC used by Sorci *et al.* contains a short, non-cleavable His×6 purification tag, while the uncleaved purification tag used in these studies is 25 amino acids long. It was reported that in some cases the presence of the purification tag may significantly affect protein activity and other functional properties (8, 9). Assuming that the purification tag impacts the activity of NadC for the kinetic studies it will be necessary to produce a recombinant protein without any additional amino acids, or with a short purification tag as described previously (1).

NadCDE Activity Assay

Despite the fact that the NadC activity was impaired, it was possible to show that NadC, NadD and NadE are able to produce NAD^+. The reaction mixture containing all QSEs and supplemented with $MgCl_2$, NH_4Cl, ATP, phosphoribosyl pyrophosphate (PRPP) and quinolinic acid was shown to produce NAD^+. This result prompted us to investigate whether the enzymes are present in solution as separate molecular entities or whether they

form higher order molecular assemblies. Each of the isolated QSEs were tested on a gel, in non-denaturing conditions, to check whether such high order molecular assemblies are formed. The results of these experiments clearly indicate that while all NadC, NadD and NadE are present in solution in an oligomeric form, the formation of binary or tertiary complexes between these enzymes does not take place in conditions that were used for the experiment (Figure 2.4).

THERMAL STABILITY

Differential scanning fluorimetry studies have shown that native spNadC is more stable than the mutant construct. SpNadC remained most stable in conditions at high salinity (1 M NaCl) and high pH (pH 9.5) and remained stable until the experiment temperature reached 60 °C (Figure 2.6). $SpNadC_{\Delta 69A}$ was most stable in high salinity (1 M NaCl), slightly alkaline (pH 7.5-8) conditions and remained stable up to a temperature of 58 °C (Figure 2.7). It should be noted that while the numerical difference between the melting temperatures of each protein is only a difference of 2 °C, while the number of conditions that were identified to provide stability up to ≥ 56 °C varied greatly. Native spNadC results showed 18 conditions (mostly 250 mM- 1M NaCl, pH 7.0-9.5) with melting point temperatures ≥ 56 °C, while $spNadC_{\Delta 69A}$ only showed stability in only two conditions within that range.

This cost-effective screening technique can also be used for identification of small molecular compounds that promote thermal stability of a protein (10, 18). Use of this approach resulted in identification of conditions in which protein was more stable, and such conditions were used during protein purification and for protein storage. The dialysis buffer

conditions that were chosen to store spNadC were: 150 mM NaCl and 0.05 M Tris (pH 7.5).

CRYSTALLIZATION

Native NadC and $NadC_{\Delta 69A}$ crystals were identified within 48 hours of crystallization setup. Crystals were identified in a variety of conditions including Index Screen (Hampton Research, Aliso Viejo CA) condition numbers 18, 19, 26, 67, and 81, Wizard Classic I (Rigaku, Bainbridge Island, WA) conditions 18, 38, and 48, and Wizard Classic II (Rigaku, Bainbridge Island, WA) conditions 15 and 48 (Table 2.1). While crystal formation was fast and plentiful, obtaining crystals that diffracted to a resolution higher than 4.0 Å was very challenging. Despite the fact that the crystals appeared from various solutions their morphology was always the same. The most promising condition (Index screen condition # 67: 0.2 M ammonium sulfate, 0.1 M BIS-TRIS (pH 6.5), 25% w/v PEG 3350) was optimized to improve crystal quality. Optimization eventually led to producing crystals that were from 5 to 250 microns in size. The biggest improvement in crystallization was achieved after use of the Hampton Research Additive screen (generous gift from Dr. James Sodetz, University of South Carolina) and identification of NaBr as the compound that improved crystal quality. The best crystals were grown from the solution composed of 0.48 M ammonium sulfate, 0.1 M BIS-TRIS (pH 6.5), 20 % w/v PEG 5000 and 1.25 mM NaBr. It was also identified that the NadC and $NadC_{\Delta 69A}$ crystal morphology was temperature dependent. At room temperature the crystals were cube-shaped, while at 4 ºC the crystals adopted an elongated shape (Figure 2.7). However, the change of the crystal morphology did not have any effect on diffraction quality. The largest crystals obtained were from a 1.1 M ammonium tartrate condition that produced crystals

that were up to 500 microns. Unfortunately, the large volume of the produced crystals did not correlate with improved diffraction quality. The most notable results were obtained while conducting *in situ* proteolysis experiments (11). In these experiments, a protease, trypsin specifically, was added in nanogram quantities to our protein stock solution directly prior to crystal screening preparations. The goal of this procedure was to minimize the entropy of the protein surface by cleaving the loop structures. As a result, we hoped to improve the probability of crystallization. The result from these experiments also promoted an increase in crystals size to approximately 500 microns. Additionally, *in situ* proteolysis crystals were observed to have a "pitted" center (Figure 2.7). The direct cause of this morphology is yet to be determined.

Analysis of NadC Sequence

SpNadC contains 299 residues in each chain of the hexameric structure (1794 total residues). The NCBI Protein BLASTsearch against homologous proteins found in the PDB reveals that this each chain is composed of 2 regions: the dimer interface and the active site (5). These results also include predicted residues of the active site (residues in **bold** are highly conserved): **T139-K141**, **H163, R164, D224, S247, G248** and dimer interface: S31, T32, **E105, R106**, **T142**, L166, S167, M170, **K173**, N175, **H176**, A179, and A190 (5).Conservation of the NadC sequences is shown in Figure 2.8.

Using the Database of Prokaryotic operons (DOOR^2) it was identified that the QSPs are not found on the same operon, as it observed in *Mycobacterium tuberculosis* (12). Database results indicate that, in reference to the origin of replication, the *nadC* reading

frame is the farthest upstream of the QSPs. Results from DOOR[2] show *nadC* at position 180951 to 181823 within the genome.

Structural Analysis

Diffraction results were first obtained from spNadC$_{\Delta 69A}$ crystals grown in the following conditions: (1) 1M ammonium phosphate, 0.1 M sodium citrate tribasic/citric acid (pH 5.5), 0.2 M sodium chloride (Wizard Classic II, #33, Rigaku Bainbridge Island, WA), and (2) 20% w/v PEG 6000, 0.1 M BIS-TRIS (pH 6.4), 0.48 ammonium sulfate, 1.25 mM sodium bromide. The crystal obtained using condition (1) belong to the cubic system and P23 space group, while the conditions (2) lead to orthorhombic crystals with $C222_1$ symmetry. The orthorhombic crystals were obtained through optimization of the Hampton Research Index screen condition # 67. The native NadC crystals were grown from solution containing from 0.48 M ammonium sulfate, 25% w/v PEG 6000 and 0.1 M BIS-TRIS (pH 6.5). The cubic form contained four protein chains per asymmetric unit and the orthorhombic form contained six protein chains per asymmetric unit. Data collection and processing details are outlined in Table 2.2. Structure solution and refinement details are outlined in Table 2.3.

Overall Structure

The first structure to be determined was the structure of the deletion mutant (spNadC$_{\Delta 69A}$), and the presence of the deletion was discovered during model building. This discovery prompted us to produce the native protein (spNadC) with the corrected sequence. All determined spNadC structures revealed the presence of hexameric assemblies in the crystals. This is in agreement with gel filtration, dynamic light scattering, and native gel

results indicating that spNadC is a hexameric structure in solution. It is assumed the hexamer observed in the crystal structure is the same hexamer found in solution. The identified oligomeric structure may be treated as a "trimer of dimers". The proposed substrate binding site is formed by two protein chains that compose the dimer. The area of interaction between chains forming the dimer is very large and is calculated to be 2623 $Å^2$, as calculated by PDBePISA (13).

The analysis of the $NadC_{\Delta 69A}$ structure revealed that the protein fragment, which should contain the missing alanine residue, is facing the opening located in the middle of the hexamer (Figure 2.10). This protein fragment does not correspond to any secondary structural element stabilized by hydrogen bonds, and the deletion does not impact the overall structure of the protein. However, the thermal stability of the mutant is decreased in comparison with the native protein. It is possible that the shortening of the loop region containing A69 results in a strain of the protein backbone that leads to the decreased mutated protein stability. It is also possible that the destabilization of the protein is related to the fact that omission of the A69 brings together two acidic and negatively charged residues (D68 and E70). The lack of any significant structural differences between the native and the mutated versions of spNadC and the relatively remote location of the deletion with respect to the active site explains why that both versions of protein display very similar enzymatic activity.

ACTIVE SITE

Structural comparisons of spNadC with seNadC (PDB code: 1QAP) bound to QA show that residues T139, R141, K141, H163, and R164 of spNadC are identical with the

residues of the stNadC structure. (Figure 2.12). Comparisons with the mtNadC (PDB code: 1QPR) bound to substrate analogs, phtalic acid (PHT) (QA analog) and PPC (PRPP analog) also indicates that the residues involved in QA and PRPP binding are highly conserved not only in terms of sequence, but also in their structure (Figure 2.13). Sharma *et al.* suggested that PRPP is bound to two Mg^{2+} ions which are considered responsible for stabilizing the pyrophosphate region of PRPP (14). However, to date there have been no structures determined with PRPP bound.

In mtNadC residues R105, R139 and R162 (which correspond to R106, R140 and R163 in spNadC) are responsible for binding of QA while D173 and R48 residues along with a series of coordinated water molecules bind to the phosphate group and metal ion on the "pyrophosphate side" of PRPP. In mtNadC the nitrogen atom from the backbone of G249, the carbonyl from A268, the nitrogen atom from G270, and the side chain of H274 are proposed to be involved in binding of the PRPP C5 phosphate group (14). No mechanistic details for this reaction are available, but it is presumed that the nitrogen of the pyridine ring of QA initiates a nucleophilic substitution (S_N1) reaction on the δ^+ C3 of PRPP to create the nucleoside bond forming NaMN and PP_i (14). Structural comparisons between spNadC and the homologs reported to the PDB also suggests that the QA and PRPP binding sites are created from residues from one chain of the dimer and not from the interaction of both chains together. Superposition of structures of homologous enzymes from *Homo sapiens*, *Salmonella enterica*, and *Mycobacterium tuberculosis* also indicates very similar substrate binding locations.

Conclusions and Future Directions

The crystal structure of spNadC was determined at 2.85Å resolution. The protein forms a hexameric assembly that is present in both solution and crystal state. It is possible that the N-terminal purification tag affects protein activity, as well as the quality of the crystals. Therefore, the next step in the project will involve generation of a protein construct with the purification tag located on the C-terminus. Moreover, lysine residues, on the protein surface, will be mutated to alaine to increase the chance of obtaining the high quality crystals diffracting to higher resolution. Such crystals are necessary for elucidation of details of interactions between spNadC and its substrates. Polyhistidine tag removal may also improve enzymatic activity of the enzyme; with the goal of observing Michaelis-Menten kinetics in increasing QA and PRPP conditions, as it was observed in the case of homologous proteins (2, 7). However, it cannot be excluded that the reaction mechanism of spNadC may be different from those reported for the homologous proteins, as the sequence identity between spNadC and the proteins that were characterized in terms of their enzymatic activity is approximately 40%.

2.3 Materials and Methods

Protein Production and Crystallization

Genomic information on the QSEs were obtained from the Uniprot (15). Synthetic genes with optimized codons were produced by DNA 2.0 (Menlo Park, CA) and introduced into a vector bearing kanamycin-resistance pJexpress411. Each of the QSEs were designed to contain an N-terminal purification tag (MHHHHHHSSGVDLGTENLYFQS↓GSG). The purification tag includes a TEV-cleavage site (marked with an arrow). Each of the plasmids

were introduced into competent BL21 (DE3) *E. coli* (New England Biolabs, Ipswich, MA), and the cells were made competent using calcium chloride (17). The plasmids were introduced using the heat/ cold shock method for bacterial transformation (18). Upon transformation, the cells were grown on Luria Broth (LB) agar plates supplemented with kanamycin, for 12 hours at 37 °C. Observations made the following day identified greater than 100 pearl-white colonies. A single colony was transferred and cultured in 5 mL of kanamycin-infused LB broth for 12-16 hours at 37 °C. After the 12-16 hour incubation, a portion of the cultures were stored in a 40 % glycerol/LB media solution and kept at -80 °C. The remaining culture was used for transformation confirmation by DNA sequencing. As it was previously mentioned, during model building it was discovered that the synthesis of nadC gene resulted in omission of a codon and production of the spNadC$_{\Delta 69A}$ deletion mutant. The gene was re-synthetized, to insert the alanine codon into the DNA sequence, by Genscript (Piscataway, NJ).

Overnight preparatory cultures of transformed BL21 (DE3) *E. coli* (New England Biolabs, Ipswich, MA) were transferred into 1L of LB media, and allowed to rotate, at 250 RPM, for approximately 2.5 hours at 37 °C. When optical density (OD_{600}) was achieved, between 0.8 and 1.0, 200 µL of 1 M IPTG was added to promote protein over-expression as outlined by Sorci *et al.* (1). The induced culture temperature was dropped to 18 °C and allowed to rotate at 250 rpm for 12-16 hours (1).

Upon completion of the 12-16 hour culture, the culture media was spun at 10524 *g* for 20 minutes at 4 °C using the Beckman Coulter Avanti J265 XPI centrifuge (Brea, CA).The resulting cell pellet was weighed and stored at -80 °C. In preparation for isolation, the cell pellet was re-solubilized in buffer containing: 500 mM NaCl, 150 mM Tris buffer

(pH 7.5), 5 mM 2-mercaptoethanol, and 1x concentration of protease inhibitors (Thermo Scientific, Waltham, MA). The cells were lysed, on ice, using the Branson 450 Sonifier. Each cell pellet was sonicated in 10, 30 second, cycles, which included a 1 minute break between cycles. The cell lysate was separated into soluble and insoluble fractions using the Beckman Allegra X30 R centrifuge, at 3849 *g*, for 20 minutes. The soluble fraction was decanted and centrifuged again using the Beckman Avanti J265 XPI centrifuge at 32264 *g* for 20 min, at 4 °C. The soluble fraction was used for protein isolation, using the immobilized metal affinity chromatography and Ni-NTA resin (Qiagen, Hilden, DE), while the insoluble fraction was kept at -80 °C for analysis. The protein was eluted from the column using an imidazole gradient of 5 mM, 60 mM, 250 mM, and 500 mM. Purity of protein samples was confirmed by SDS-PAGE. The eluted protein solution was allowed to dialyze overnight, at 4 °C, in buffer containing: 50 mM Tris, 150 mM NaCl, and 5 mM 2-mercaptoethanol. The dialyzed protein solution was further purified by size exclusion chromatography. All size exclusion experiments were conducted on the AKTA™ Pure (General Electric, Piscataway, NJ) using the Superdex 200 HiLoad 16/60 column. Buffer conditions for these experiments were the same as for the dialysis buffer. The fractions representative of the protein of interest were concentrated, divided into 1 mL aliquots, and stored at -80 °C.

All initial crystallization experiments were performed using the sitting-drop vapor-diffusion method. The initial screening utilized the 96 well Original Intelli 96-2 plates (Art Robbins Instruments, Sunnyvale, CA). The well contained 80 µL of screening solution, and the drops were formed by mixing of 1 µL of protein solution and 1 µL of screening solution. Several commercial crystallization screens, Wizard Classic 1-4 (Molecular

Dimensions/Rigaku, Altamonte Springs, FL) and Index (Hampton Research, Aliso Viejo CA) for example, were used in the initial search of crystallization conditions.

Optimization experiments used both sitting and hanging drop techniques. Sitting-drop experiments used the 24-well Cryschem sitting-drop plate (Hampton Research, Aliso Viejo CA). Each well contained 500 µL of crystallization solution, and the sitting-drop contained 3 µL of 2 mg/mL protein and 3 µL of well solution. Hanging-drop experiments utilized the Qiagen EasyXtal 15 well tool (Hilde, DE). The well contained 150 µL of crystallization solution, and the screw cap contained 2 drops, 6 µL each, in the same solution mixture as in the sitting-drop experiment Most of the crystallization experiments were conducted at 25 °C, with fewer experiments conducted at 18 °C and 4 °C.

Dynamic Light Scattering

Dynamic Light Scattering (DLS) was used to characterize a quaternary structure of protein in solution. For DLS experiments, 2 mL of 2 mg/mL QSE was dialyzed for 12 hours in the dialysis solution described above. After dialysis, the QSEs were diluted to 0.2 mg/mL, using the dialysis buffer, and the sample was analyzed using the experimental procedure for instrumental analysis, as outlined by protocol prepared by Malvern (Worcestershire, UK) for the Zetaseizer ZS90.

Differential Scanning Fluorimetry

Differential Scanning Fluorimetry (DSF) was used for the determination of protein thermal stability as well as the determination of the small molecular compounds influence on protein stability (19). SYPRO-Orange is an additive that produces fluorescence when bound to hydrophobic residues during protein denaturation (19, 20). In this experiment,

proteins are heated incrementally, and the determination of their thermal stability is indicated by the temperature at which they fluoresce.

Protein stability was initially checked by testing 96 different buffer conditions using an in-house pH and salinity screen. The screen was designed to increase pH from 4.0 to 9.5 and increase NaCl concentrations from 0.0 to 1.0 M using a 96-well arrangement (19). A chart displaying the buffer strategy is outlined in Figure 2.13. Sypro-Orange (Life Technologies, Grand Island, NY) was first diluted 1:1000 in 1 mL of the QSE stock, at 2 mg/mL, and the protein solution was placed on ice. Ninety-six buffer solutions, increasing in pH and NaCl concentration, were then added to their respective wells using the Biorad (Hercules, CA) Hardshell 96-well PCR plate. Figure 2.14 provides a schematic describing buffer conditions for each well. Using the BioRad (Hercules, CA) CFX96 real-time PCR machine, a computer protocol was designed to vary the temperature of each well by 2 ºC/ min from 30 ºC to 90 ºC. Melting temperatures for the QSEs were determined by fluorescence. Derivative results for protein melting points were calculated by the Bio-Rad CFX96 software package. This protocol was adapted by Dr. Nicholas Mank (unpublished) from the original work of Niesen *et al.* (19). The stability of QSEs was also tested in the presence of various small molecular compounds. Experiment preparation was the same as described above. All small molecule screening conditions were made soluble in aqueous solution and were buffered with Tris to pH 7.5. A list of all 384 conditions is shown in Appendix A.

KINETICS

The enzymatic activity of spNadC was determined by a method outlined in the study of a homologus quinolinate phosphoribosyl transferase in *Salmonella tyhphimurium* (7). Briefly, activity reaction conditions, at a final volume of 1 mL, contained 50 mM HEPES (pH 7.5), 6 mM $MgCl_2$, 200 µM PRPP, 300 µM QA, and 20 µg of NadC. The reaction time was 10 min per experiment. Reaction conditions for kinetic experiments which involved testing the effect of increasing PRPP on the activity of spNadC maintained the same activity assay concentrations, described above, with increasing amounts of PRPP at the following concentrations: 200 µM, 400 µM, 600 µM, 800 µM, 1 mM, 1.2 mM, 1.4 mM, and 1.6 mM. The above experiment was repeated with increasing amounts of QA at the following concentrations: 2 µM, 4 µM, 6 µM, 8 µM, 10 µM, 12 µM, 14 µM, 16 µM, 18 µM, and 20 µM. NaMN detection was measured using the Nanodrop 2000c (Thermo Scientific, Waltman, MA) at 266 nm. Each experiment was done in triplicate, at 1 mL, using a quartz cuvette, and NaMN biosynthesis was detected at 266 nm.

In a separate assay, biosynthesis of NAD^+ was confirmed, at 340 nm, with the use of all of the QSEs in one reaction solution. The reaction was conducted at 1 mL in a quartz cuvette. The reaction solution contained: 50 mM HEPES (pH 7.5), 6 mM $MgCl_2$, 20 mM KCl, 10 mM NH_4Cl, 2 mM ATP, 0.2 mM phosphoribosyl pyrophosphate (PRPP), 0.36 mM quinolinic acid (QA), 2 µg bovine serum albumin (BSA), 1 % ethanol, 4 µg NadC, 4 µg NadC, 4 µg NadE, 10 µg alcohol dehydrogenase (ADH). The activity assay was run for 10 min. The success of this assay lead us to attempt a supplemental assay to determine the effect of increasing NadC on the rate of NAD^+ formation, although no conclusive results have been obtained to date.

Data Collection and Processing

Prior to the diffraction experiments spNadC and spNadC$_{\Delta 69A}$ crystals were cryo-cooled. 10 % glycerol was used to cryo-protect the crystal grown from ammonium phosphate, and no cryoprotectant was used for other crystals. Data was collected at the Advanced Photon Source at Argonne National Laboratory in Lemont, IL. The data from the NadC$_{\Delta 69}$ crystals were collected at Sector 22 belonging to the Southeast Regional Collaborative Access Team (SER-CAT); beamlines 22 BM and 22 ID. Data for native spNadC was collected at Sector 19 belonging to the Structural Biology Center; beamline 19 BM. All data collection was performed at 100 K. Data processing was performed using HKL-2000 (21) and the details of the diffraction experiments are shown in Table 2.2.

Structure Determination and Validation

For each of the spNadC and spNadC$_{\Delta 69A}$ structures HKL-3000 and MOLREP (21, 22) were used for structure determination. Both of the spNadC$_{\Delta 69A}$ structures were solved prior to the spNadC structure and molecular replacement was used for phasing. The P23 spNadC$_{\Delta 69A}$ structure was solved first using the *Homo sapiens* NadC homolog (PDB code: 4KWV) as a starting model. The model of spNadC$_{\Delta 69A}$ cubic form was used as a starting point for determination of the orthorhombic $C222_1$ form. Determination of native spNadC structure was done by Fourier synthesis using the spNadC$_{\Delta 69A}$ $C222_1$ structure to calculate phases. All structures were refined using HKL-3000 and Refmac (23), and models were updated and validated using COOT (24). Final validation was done using Molprobity (25) and CheckMyMetal (26). All of the spNadC and spNadC$_{\Delta 69A}$ structures and structure factors were deposited into the Protein Data Bank (PDB) and were given the accession

codes 5HUL (spNadC$_{\Delta 69A}$ - P23), 5HUO (spNadC$_{\Delta 69A}$ - C222$_1$), and 5HUP (native spNadC). Summary information on model refinement and validation is included in Table 2.3.

Other Computational Methods

Other Computational methods include: the DALI database for protein structural comparison studies (27); Xtalpred, for predicative studies on the likelihood of crystallization of all of the QSPs (28); Protparam, for the determination of predicted chemical parameters for all of the QSPs (29). PHYRE2 and SwissModel were used to make predictions on the structural parameters for each of the QSPs through structural homology calculations (30, 31). Pymol was used for structure structural studies and cartoon and surface representations throughout this dissertation (32). Clustal Omega and ESPript was employed for sequence alignment and secondary structure studies (33, 34). PDBePISA was applied for the calculation of structural and thermodynamic parameters for each of the QSPs (35). Consurf was used for the determination of residue conservation for all of the QSPS, as well as provided structural representations for this conservation for display in Pymol (36).

2.4 TABLES AND FIGURES

Table 2.1: Crystallization conditions for spNadC and spNadC$_{\Delta 69A}$ identified during an initial screening.

Screen	Condition #	Condition contents
Index	18	0.49 M sodium phosphate monobasic monohydrate, 0.91 M potassium phosphate dibasic (pH 6.9)
	19	0.056 M sodium phosphate monobasic monohydrate, 1.344 M potassium phosphate dibasic (pH 6.9)
	26	1.1 M ammonium tartrate (pH 7.0)
	67	0.2 ammonium sulfate, 25 % w/v PEG 3350, 0.1 M BIS-TRIS (pH 6.5)
	81	0.2 M ammonium acetate, 25 % w/v PEG 3350, 0.1 M HEPES (pH 7.5)
Wizard Classic I	18	1.0 M potassium sodium tartrate, 0.2 M sodium chloride, 0.1 M imidazole (pH 8.0)
	38	0.2 M lithium sulfate, 1.0 M potassium sodium tartrate, 0.1 M CHES (pH 9.5)
	48	20 % w/v PEG 1000, 0.2 M zinc acetate, 0.1 M sodium acetate/ acetic acid (pH 4.5)
Wizard Classic II	15	1.26 M ammonium sulfate, 0.1 M HEPES (pH 7.5)
	48	1.0 M potassium sodium tartrate, 0.1 M MES (pH 6.0)

Table 2.2: Summary of spNadC and spNadC$_{\Delta 69A}$ data collection and processing.

Protein	spNadC$_{\Delta 69A}$	spNadC$_{\Delta 69A}$	spNadC
PDB accession code	5HUL	5HUO	5HUP
Diffraction source	APS (21 ID)	APS (22 BM)	APS (19 BM)
Wavelength (Å)	1.000	1.000	1.000
Detector	Rayonix MX300HS	Rayonix MAR 225	ADSC Q210r
Crystal-to-detector distance (mm)	380	280	270
Rotation range per image (°)	0.5	1.0	0.5
Total rotation range (°)	100	200	100
Exposure time per image (s)	1	1	1
Space group	P23	$C222_1$	$C222_1$
a, *b*, *c* (Å)	179.6, 179.6, 179.6	106.2,186.3, 221.9	108.2, 189.0, 222.4
Resolution range (Å)	40- 2.85 (2.95- 2.85)	40- 2.80 (2.85- 2.80)	50.00- 3.45 (3.46- 3.40)
σ Cutoff	-3 σ	-3 σ	-3 σ
Total No. of reflections	44917 (2166)	54566 (2658)	26521 (1389)
Completeness (%)	99.8 (99.5)	100 (100)	90.3 (95.2)
Redundancy	4.7 (4.6)	8.4 (8.5)	3.8 (3.9)
$\langle I/\sigma(I)\rangle$	24.1 (2.0)	14.6 (2.3)	7.9 (1.8)
$R_{r.i.m.}$	0.129 (0.990)	0.162 (0.958)	0.140 (0.790)
$R_{p.i.m.}$	0.055 (0.434)	0.056 (0.331)	0.071 (0.389)
Overall *B* factor from Wilson plot (Å^2)	84.6	64.6	45.3
CC ½	(0.573)	(0.791)	(0.729)

Table 2.3. Summary of spNadC and spNadC$_{\Delta 69A}$ structure refinement statistics.

Protein	NadC$_{\Delta 69A}$	NadC$_{\Delta 69A}$	Native NadC
Resolution range (Å)	39.23- 2.86 (2.92- 2.85)	37.01- 2.80 (2.87- 2.80)	50- 3.45 (3.54- 3.45)
Completeness (%)	99.72 (99.6)	99.92 (100)	82.40 (72)
σ Cut-off	0	0	0
No. of reflections, working set	42638	51691	23687
No. of reflections, test	2268	2768	1271
Final R_{cryst}	0.222	0.220	0.303
Final R_{free}	0.246	0.258	0.348
No. of non-H atoms	8629	13034	12565
Protein	8517	12798	12454
Ligand	16	20	17
Solvent	96	216	94
R.m.s deviations			
Bonds (Å)	0.012	0.012	0.011
Angles (°)	80	53	84
Average B factors (Å^2)			
Protein	81.5	54.0	88.3
Ligand	86.1	85.1	85.1
Ramachandran plot			
Most favored (%)	96.6	96.7	93.5
Allowed (%)	100	99	99.5
Visible Residues			
Chain A	6- 288	6- 288	6- 288
Chain B	6- 288	6- 288	6- 288
Chain C	6- 288	6- 288	8- 289
Chain D	5-288		
Chain E		6- 288	6- 289
Chain F			5-289
Chain H			6-289

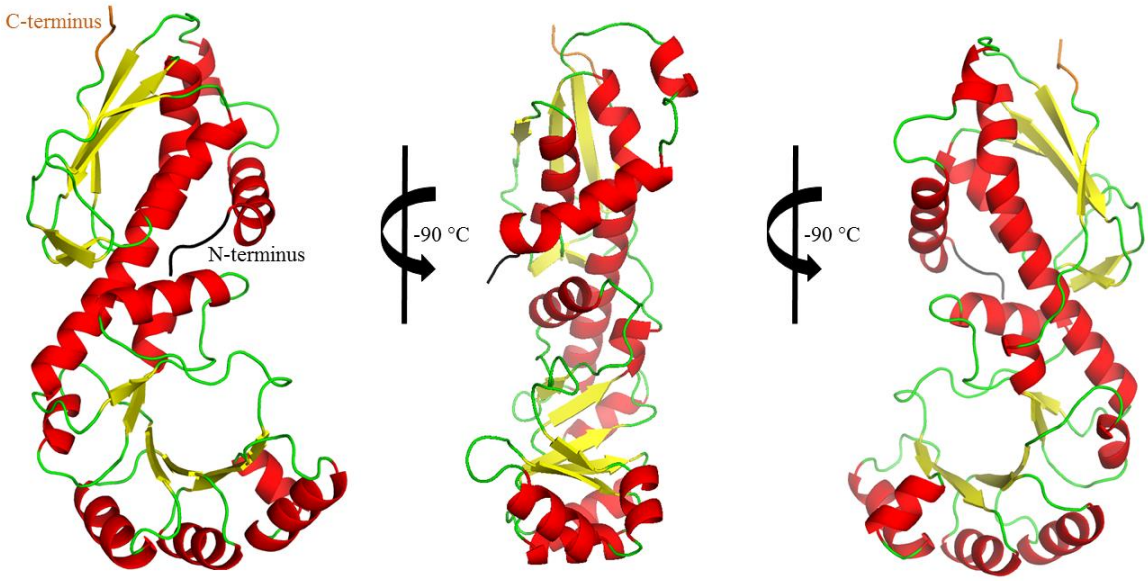


Figure 2.1: The spNadC monomer. This is a cartoon representation of a single chain of spNadC shown in three different orientations. The N-terminus is shown in black, the C-terminus is colored orange.

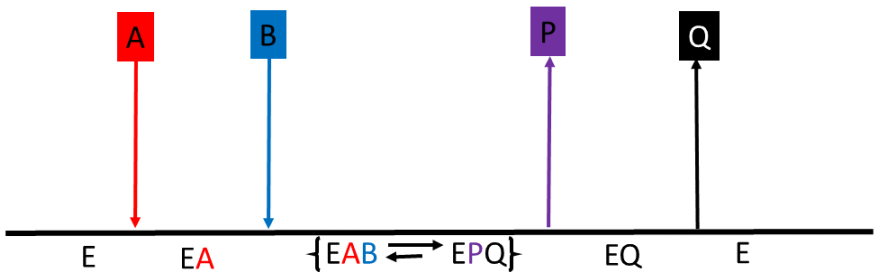


Figure 2.2: Cleland diagram describing ordered-sequential reaction. This diagram shows an overview of an ordered-sequential reaction proposed for spNadC as based on Cao *et al.* (7, 35). This schema is shown using the Cleland notation (35). "A" corresponds to PRPP; "B" corresponds to QA; "P" stands for NaMN (released); "Q" describes CO_2 and PP_i (released).

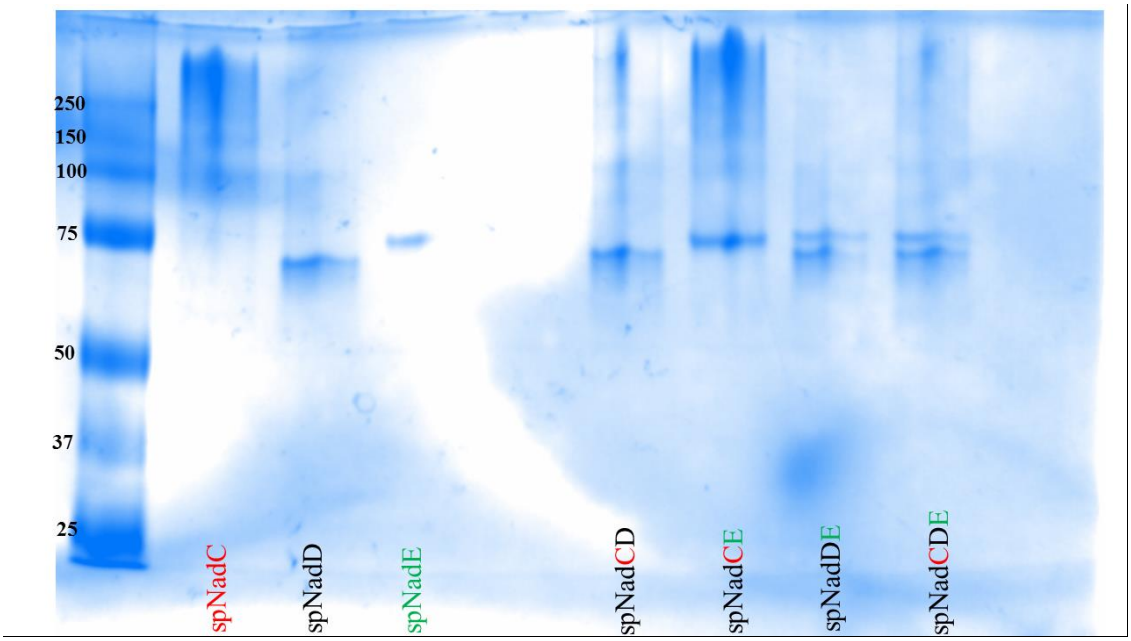
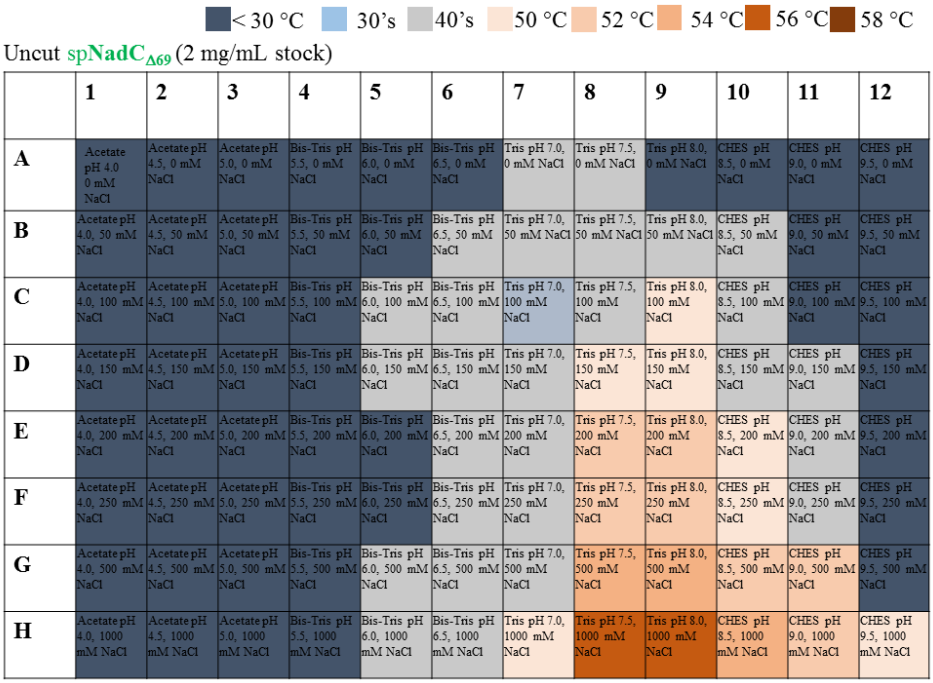


Figure 2.3: Native gel electrophoresis of spNadC, spNadD, and spNadE. These results are of the native gel electrophoresis for the experiments testing formation of higher order molecular assemblies between NadC, NadD and/or NadE. No complex formation was observed.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Acetate pH 4.0 0 mM NaCl	Acetate pH 4.5, 0 mM NaCl	Acetate pH 5.0, 0 mM NaCl	Bis-Tris pH 5.5, 0 mM NaCl	Bis-Tris pH 6.0, 0 mM NaCl	Bis-Tris pH 6.5, 0 mM NaCl	Tris pH 7.0, 0 mM NaCl	Tris pH 7.5, 0 mM NaCl	Tris pH 8.0, 0 mM NaCl	CHES pH 8.5, 0 mM NaCl	CHES pH 9.0, 0 mM NaCl	CHES pH 9.5, 0 mM NaCl
B	Acetate pH 4.0, 50 mM NaCl	Acetate pH 4.5, 50 mM NaCl	Acetate pH 5.0, 50 mM NaCl	Bis-Tris pH 5.5, 50 mM NaCl	Bis-Tris pH 6.0, 50 mM NaCl	Bis-Tris pH 6.5, 50 mM NaCl	Tris pH 7.0, 50 mM NaCl	Tris pH 7.5, 50 mM NaCl	Tris pH 8.0, 50 mM NaCl	CHES pH 8.5, 50 mM NaCl	CHES pH 9.0, 50 mM NaCl	CHES pH 9.5, 50 mM NaCl
C	Acetate pH 4.0, 100 mM NaCl	Acetate pH 4.5, 100 mM NaCl	Acetate pH 5.0, 100 mM NaCl	Bis-Tris pH 5.5, 100 mM NaCl	Bis-Tris pH 6.0, 100 mM NaCl	Bis-Tris pH 6.5, 100 mM NaCl	Tris pH 7.0, 100 mM NaCl	Tris pH 7.5, 100 mM NaCl	Tris pH 8.0, 100 mM NaCl	CHES pH 8.5, 100 mM NaCl	CHES pH 9.0, 100 mM NaCl	CHES pH 9.5, 100 mM NaCl
D	Acetate pH 4.0, 150 mM NaCl	Acetate pH 4.5, 150 mM NaCl	Acetate pH 5.0, 150 mM NaCl	Bis-Tris pH 5.5, 150 mM NaCl	Bis-Tris pH 6.0, 150 mM NaCl	Bis-Tris pH 6.5, 150 mM NaCl	Tris pH 7.0, 150 mM NaCl	Tris pH 7.5, 150 mM NaCl	Tris pH 8.0, 150 mM NaCl	CHES pH 8.5, 150 mM NaCl	CHES pH 9.0, 150 mM NaCl	CHES pH 9.5, 150 mM NaCl
E	Acetate pH 4.0, 200 mM NaCl	Acetate pH 4.5, 200 mM NaCl	Acetate pH 5.0, 200 mM NaCl	Bis-Tris pH 5.5, 200 mM NaCl	Bis-Tris pH 6.0, 200 mM NaCl	Bis-Tris pH 6.5, 200 mM NaCl	Tris pH 7.0, 200 mM NaCl	Tris pH 7.5, 200 mM NaCl	Tris pH 8.0, 200 mM NaCl	CHES pH 8.5, 200 mM NaCl	CHES pH 9.0, 200 mM NaCl	CHES pH 9.5, 200 mM NaCl
F	Acetate pH 4.0, 250 mM NaCl	Acetate pH 4.5, 250 mM NaCl	Acetate pH 5.0, 250 mM NaCl	Bis-Tris pH 5.5, 250 mM NaCl	Bis-Tris pH 6.0, 250 mM NaCl	Bis-Tris pH 6.5, 250 mM NaCl	Tris pH 7.0, 250 mM NaCl	Tris pH 7.5, 250 mM NaCl	Tris pH 8.0, 250 mM NaCl	CHES pH 8.5, 250 mM NaCl	CHES pH 9.0, 250 mM NaCl	CHES pH 9.5, 250 mM NaCl
G	Acetate pH 4.0, 500 mM NaCl	Acetate pH 4.5, 500 mM NaCl	Acetate pH 5.0, 500 mM NaCl	Bis-Tris pH 5.5, 500 mM NaCl	Bis-Tris pH 6.0, 500 mM NaCl	Bis-Tris pH 6.5, 500 mM NaCl	Tris pH 7.0, 500 mM NaCl	Tris pH 7.5, 500 mM NaCl	Tris pH 8.0, 500 mM NaCl	CHES pH 8.5, 500 mM NaCl	CHES pH 9.0, 500 mM NaCl	CHES pH 9.5, 500 mM NaCl
H	Acetate pH 4.0, 1000 mM NaCl	Acetate pH 4.5, 1000 mM NaCl	Acetate pH 5.0, 1000 mM NaCl	Bis-Tris pH 5.5, 1000 mM NaCl	Bis-Tris pH 6.0, 1000 mM NaCl	Bis-Tris pH 6.5, 1000 mM NaCl	Tris pH 7.0, 1000 mM NaCl	Tris pH 7.5, 1000 mM NaCl	Tris pH 8.0, 1000 mM NaCl	CHES pH 8.5, 1000 mM NaCl	CHES pH 9.0, 1000 mM NaCl	CHES pH 9.5, 1000 mM NaCl

Figure 2.4: DSF experimental outline. This outline displays the buffer conditions within the 96-well plate for DSF/ Thermal stability experiments. These conditions were used for testing of protein thermal stability in conditions with various pH and NaCl concentrations.



	1	2	3	4	5	6	7	8	9	10	11	12
A	Acetate pH 4.0 0 mM NaCl	Acetate pH 4.5, 0 mM NaCl	Acetate pH 5.0, 0 mM NaCl	Bis-Tris pH 5.5, 0 mM NaCl	Bis-Tris pH 6.0, 0 mM NaCl	Bis-Tris pH 6.5, 0 mM NaCl	Tris pH 7.0, 0 mM NaCl	Tris pH 7.5, 0 mM NaCl	Tris pH 8.0, 0 mM NaCl	CHES pH 8.5, 0 mM NaCl	CHES pH 9.0, 0 mM NaCl	CHES pH 9.5, 0 mM NaCl
B	Acetate pH 4.0, 50 mM NaCl	Acetate pH 4.5, 50 mM NaCl	Acetate pH 5.0, 50 mM NaCl	Bis-Tris pH 5.5, 50 mM NaCl	Bis-Tris pH 6.0, 50 mM NaCl	Bis-Tris pH 6.5, 50 mM NaCl	Tris pH 7.0, 50 mM NaCl	Tris pH 7.5, 50 mM NaCl	Tris pH 8.0, 50 mM NaCl	CHES pH 8.5, 50 mM NaCl	CHES pH 9.0, 50 mM NaCl	CHES pH 9.5, 50 mM NaCl
C	Acetate pH 4.0, 100 mM NaCl	Acetate pH 4.5, 100 mM NaCl	Acetate pH 5.0, 100 mM NaCl	Bis-Tris pH 5.5, 100 mM NaCl	Bis-Tris pH 6.0, 100 mM NaCl	Bis-Tris pH 6.5, 100 mM NaCl	Tris pH 7.0, 100 mM NaCl	Tris pH 7.5, 100 mM NaCl	Tris pH 8.0, 100 mM NaCl	CHES pH 8.5, 100 mM NaCl	CHES pH 9.0, 100 mM NaCl	CHES pH 9.5, 100 mM NaCl
D	Acetate pH 4.0, 150 mM NaCl	Acetate pH 4.5, 150 mM NaCl	Acetate pH 5.0, 150 mM NaCl	Bis-Tris pH 5.5, 150 mM NaCl	Bis-Tris pH 6.0, 150 mM NaCl	Bis-Tris pH 6.5, 150 mM NaCl	Tris pH 7.0, 150 mM NaCl	Tris pH 7.5, 150 mM NaCl	Tris pH 8.0, 150 mM NaCl	CHES pH 8.5, 150 mM NaCl	CHES pH 9.0, 150 mM NaCl	CHES pH 9.5, 150 mM NaCl
E	Acetate pH 4.0, 200 mM NaCl	Acetate pH 4.5, 200 mM NaCl	Acetate pH 5.0, 200 mM NaCl	Bis-Tris pH 5.5, 200 mM NaCl	Bis-Tris pH 6.0, 200 mM NaCl	Bis-Tris pH 6.5, 200 mM NaCl	Tris pH 7.0, 200 mM NaCl	Tris pH 7.5, 200 mM NaCl	Tris pH 8.0, 200 mM NaCl	CHES pH 8.5, 200 mM NaCl	CHES pH 9.0, 200 mM NaCl	CHES pH 9.5, 200 mM NaCl
F	Acetate pH 4.0, 250 mM NaCl	Acetate pH 4.5, 250 mM NaCl	Acetate pH 5.0, 250 mM NaCl	Bis-Tris pH 5.5, 250 mM NaCl	Bis-Tris pH 6.0, 250 mM NaCl	Bis-Tris pH 6.5, 250 mM NaCl	Tris pH 7.0, 250 mM NaCl	Tris pH 7.5, 250 mM NaCl	Tris pH 8.0, 250 mM NaCl	CHES pH 8.5, 250 mM NaCl	CHES pH 9.0, 250 mM NaCl	CHES pH 9.5, 250 mM NaCl
G	Acetate pH 4.0, 500 mM NaCl	Acetate pH 4.5, 500 mM NaCl	Acetate pH 5.0, 500 mM NaCl	Bis-Tris pH 5.5, 500 mM NaCl	Bis-Tris pH 6.0, 500 mM NaCl	Bis-Tris pH 6.5, 500 mM NaCl	Tris pH 7.0, 500 mM NaCl	Tris pH 7.5, 500 mM NaCl	Tris pH 8.0, 500 mM NaCl	CHES pH 8.5, 500 mM NaCl	CHES pH 9.0, 500 mM NaCl	CHES pH 9.5, 500 mM NaCl
H	Acetate pH 4.0, 1000 mM NaCl	Acetate pH 4.5, 1000 mM NaCl	Acetate pH 5.0, 1000 mM NaCl	Bis-Tris pH 5.5, 1000 mM NaCl	Bis-Tris pH 6.0, 1000 mM NaCl	Bis-Tris pH 6.5, 1000 mM NaCl	Tris pH 7.0, 1000 mM NaCl	Tris pH 7.5, 1000 mM NaCl	Tris pH 8.0, 1000 mM NaCl	CHES pH 8.5, 1000 mM NaCl	CHES pH 9.0, 1000 mM NaCl	CHES pH 9.5, 1000 mM NaCl

Figure 2.5: Thermal stability results for spNadC$_{\Delta 69A}$. This chart depicts exemplary results of a single differential scanning fluorimetry experiment for spNadC$_{\Delta 69A}$.

< 30 °C 30's 40's 50 °C 52 °C 54 °C 56 °C 58 °C 60 °C

Uncut **Native spNadC** (mg/mL stock)

	1	2	3	4	5	6	7	8	9	10	11	12
A	Acetate pH 4.0 0 mM NaCl	Acetate pH 4.5, 0 mM NaCl	Acetate pH 5.0, 0 mM NaCl	Bis-Tris pH 5.5, 0 mM NaCl	Bis-Tris pH 6.0, 0 mM NaCl	Bis-Tris pH 6.5, 0 mM NaCl	Tris pH 7.0, 0 mM NaCl	Tris pH 7.5, 0 mM NaCl	Tris pH 8.0, 0 mM NaCl	CHES pH 8.5, 0 mM NaCl	CHES pH 9.0, 0 mM NaCl	CHES pH 9.5, 0 mM NaCl
B	Acetate pH 4.0, 50 mM NaCl	Acetate pH 4.5, 50 mM NaCl	Acetate pH 5.0, 50 mM NaCl	Bis-Tris pH 5.5, 50 mM NaCl	Bis-Tris pH 6.0, 50 mM NaCl	Bis-Tris pH 6.5, 50 mM NaCl	Tris pH 7.0, 50 mM NaCl	Tris pH 7.5, 50 mM NaCl	Tris pH 8.0, 50 mM NaCl	CHES pH 8.5, 50 mM NaCl	CHES pH 9.0, 50 mM NaCl	CHES pH 9.5, 50 mM NaCl
C	Acetate pH 4.0, 100 mM NaCl	Acetate pH 4.5, 100 mM NaCl	Acetate pH 5.0, 100 mM NaCl	Bis-Tris pH 5.5, 100 mM NaCl	Bis-Tris pH 6.0, 100 mM NaCl	Bis-Tris pH 6.5, 100 mM NaCl	Tris pH 7.0, 100 mM NaCl	Tris pH 7.5, 100 mM NaCl	Tris pH 8.0, 100 mM NaCl	CHES pH 8.5, 100 mM NaCl	CHES pH 9.0, 100 mM NaCl	CHES pH 9.5, 100 mM NaCl
D	Acetate pH 4.0, 150 mM NaCl	Acetate pH 4.5, 150 mM NaCl	Acetate pH 5.0, 150 mM NaCl	Bis-Tris pH 5.5, 150 mM NaCl	Bis-Tris pH 6.0, 150 mM NaCl	Bis-Tris pH 6.5, 150 mM NaCl	Tris pH 7.0, 150 mM NaCl	Tris pH 7.5, 150 mM NaCl	Tris pH 8.0, 150 mM NaCl	CHES pH 8.5, 150 mM NaCl	CHES pH 9.0, 150 mM NaCl	CHES pH 9.5, 150 mM NaCl
E	Acetate pH 4.0, 200 mM NaCl	Acetate pH 4.5, 200 mM NaCl	Acetate pH 5.0, 200 mM NaCl	Bis-Tris pH 5.5, 200 mM NaCl	Bis-Tris pH 6.0, 200 mM NaCl	Bis-Tris pH 6.5, 200 mM NaCl	Tris pH 7.0, 200 mM NaCl	Tris pH 7.5, 200 mM NaCl	Tris pH 8.0, 200 mM NaCl	CHES pH 8.5, 200 mM NaCl	CHES pH 9.0, 200 mM NaCl	CHES pH 9.5, 200 mM NaCl
F	Acetate pH 4.0, 250 mM NaCl	Acetate pH 4.5, 250 mM NaCl	Acetate pH 5.0, 250 mM NaCl	Bis-Tris pH 5.5, 250 mM NaCl	Bis-Tris pH 6.0, 250 mM NaCl	Bis-Tris pH 6.5, 250 mM NaCl	Tris pH 7.0, 250 mM NaCl	Tris pH 7.5, 250 mM NaCl	Tris pH 8.0, 250 mM NaCl	CHES pH 8.5, 250 mM NaCl	CHES pH 9.0, 250 mM NaCl	CHES pH 9.5, 250 mM NaCl
G	Acetate pH 4.0, 500 mM NaCl	Acetate pH 4.5, 500 mM NaCl	Acetate pH 5.0, 500 mM NaCl	Bis-Tris pH 5.5, 500 mM NaCl	Bis-Tris pH 6.0, 500 mM NaCl	Bis-Tris pH 6.5, 500 mM NaCl	Tris pH 7.0, 500 mM NaCl	Tris pH 7.5, 500 mM NaCl	Tris pH 8.0, 500 mM NaCl	CHES pH 8.5, 500 mM NaCl	CHES pH 9.0, 500 mM NaCl	CHES pH 9.5, 500 mM NaCl
H	Acetate pH 4.0, 1000 mM NaCl	Acetate pH 4.5, 1000 mM NaCl	Acetate pH 5.0, 1000 mM NaCl	Bis-Tris pH 5.5, 1000 mM NaCl	Bis-Tris pH 6.0, 1000 mM NaCl	Bis-Tris pH 6.5, 1000 mM NaCl	Tris pH 7.0, 1000 mM NaCl	Tris pH 7.5, 1000 mM NaCl	Tris pH 8.0, 1000 mM NaCl	CHES pH 8.5, 1000 mM NaCl	CHES pH 9.0, 1000 mM NaCl	CHES pH 9.5, 1000 mM NaCl

Figure 2.6: Thermal stability results for spNadC. This chart depicts results of a single differential scanning fluorimetry experiment for wild-type spNadC.

Figure 2.7: SpNadC crystals with various morphologies. These figures show wild-type spNadC crystals with different morphologies obtained in various conditions. The leftmost picture displays the "pitted" center identified in *in situ* proteolysis conditions, the middle figure displays the results of crystallization at 4 ºC and formation of an elongated spNadC crystal. The right figure displays the most common spNadC crystal morphology observed for crystals grown at 25 °C.

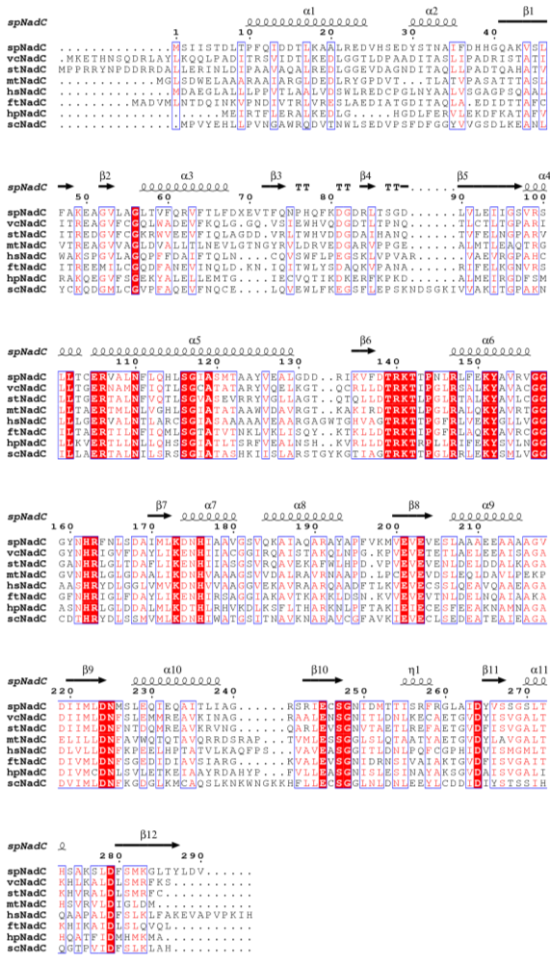


Figure 2.8: SpNadC sequence alignment. Sequence alignment for spNadC and homologous proteins which have their structures determined. The alignment was done using the ESPript 3.0 and CLUSTALW. SpNadC homologs originate from *Vibrio cholerae* (vcNadC), *Salmonella enterica* (stNadC), *Mycobacterium tuberculosis* (mtNadC), *Homo sapiens* (hsNadC), *Franciscella tularensis* (ftNadC), *Helicobacter pylori* (hpNadC), and *Saccharomyces cerevisiae* (scNadC).

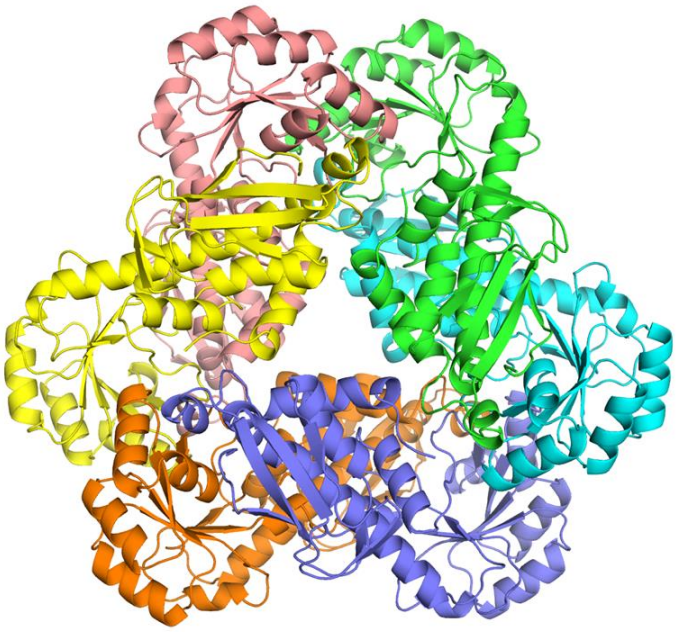


Figure 2.9: SpNadC structure. Cartoon representation of the hexameric assembly of spNadC

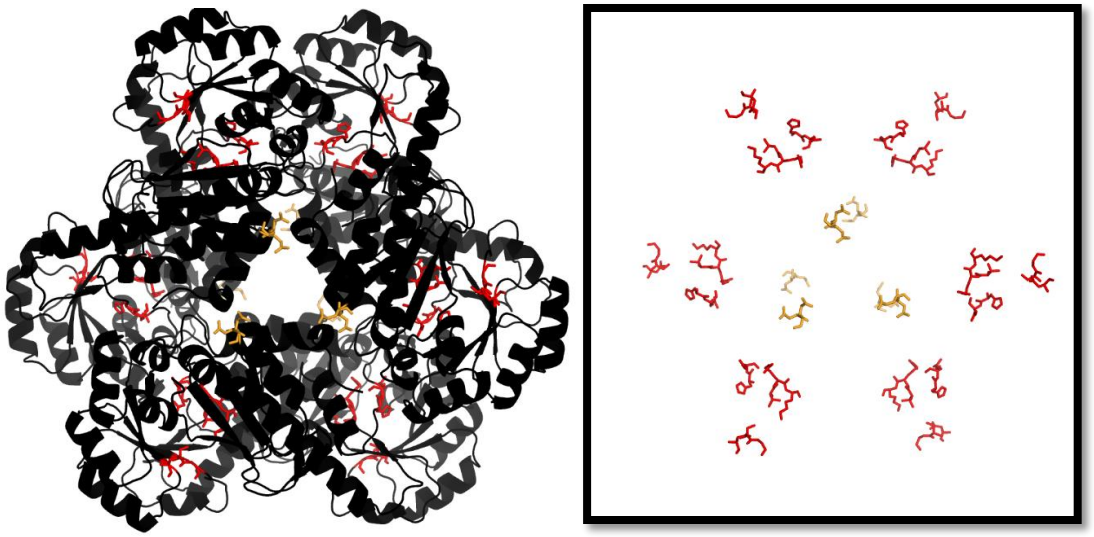


Figure 2.10: SpNadC$_{\Delta 69A}$ structure. Cartoon representations of the hexameric assembly of NadC$_{\Delta 69A}$ with (*left*) and without (*right*) secondary structure elements. Areas highlighted in yellow define the locations of D68 and E70 residues. Areas highlighted in red are proposed substrate binding sites.

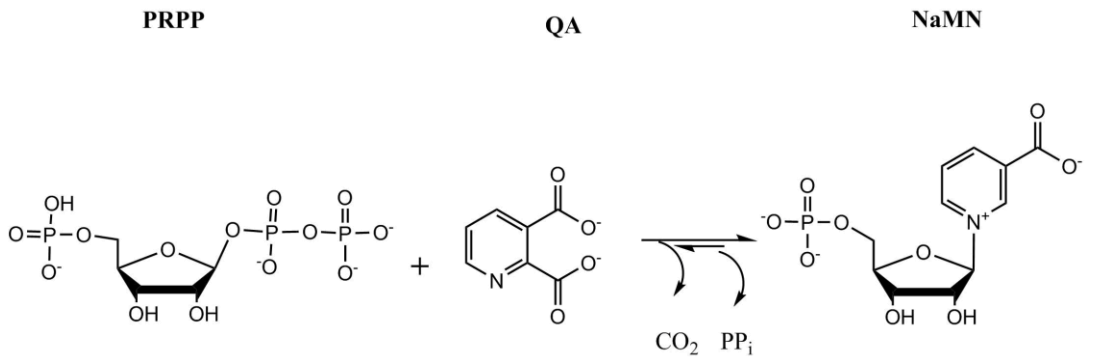


Figure 2.11: Generalized spNadC reaction. Schematic of the reaction catalyzed by spNadC.

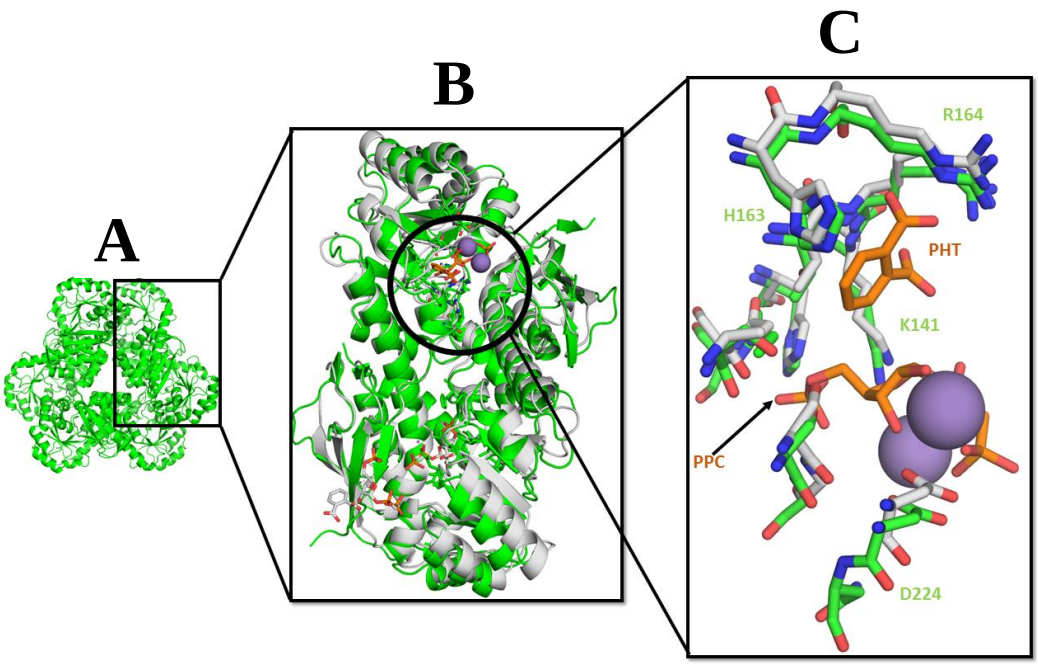


Figure 2.12: SpNadC with modeled inhibitor molecules. This cartoon representation shows structural features of spNadC. Section "A" is representative of the complete spNadC hexameric assembly; section "B" shows an spNadC dimer magnified and superposed with a homologous dimer from mtNadC (PDB code: 1QPR) to show the predicted site active site for spNadC (mtNadC is also hexamer); section "C" shows the magnified active site and displays residue similarities between each structure. The residue numbers describe the amino acids from the spNadC structure only.

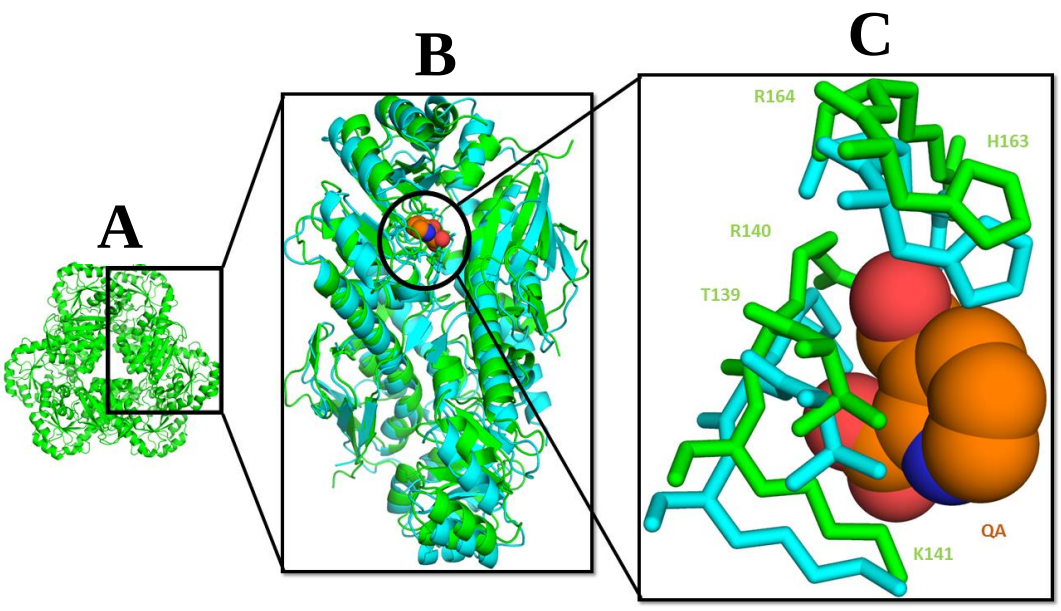


Figure 2.13: SpNadC with modeled substrate. This cartoon representation shows structural features of spNadC. Section "A" is representative of the complete spNadC hexameric assembly; section "B" shows an spNadC dimer magnified and superposed with a homologous dimer from stNadC (PDB code: 1QAP) to show the predicted site active site for spNadC (stNadC is a dimer); section "C" shows the magnified active site and displays residue similarities between each structure. The residue numbers describe the amino acids from the spNadC structure only

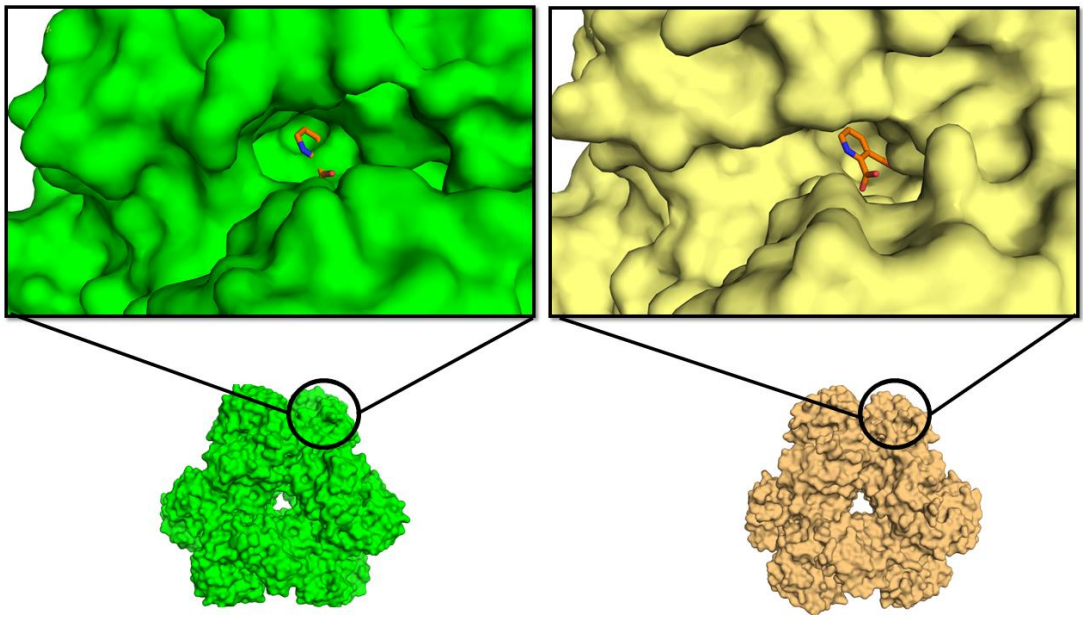


Figure 2.14: Putative spNadC QA binding pocket. This figure shows a surface representation of spNadC (left) and hsNadC (PDB code: 5AYY) (right) and the QA binding pocket. For spNadC, the position of the QA was modeled using the hsNadC structure. QA is shown in stick representation.

2.5 REFERENCES:

1. Sorci, L., I. K. Blaby, I. A. Rodionova, J. De Ingeniis, S. Tkachenko, V. De Crecy-Lagard, and A. L. Osterman. "Quinolinate Salvage and Insights for Targeting NAD Biosynthesis in Group A Streptococci." *Journal of Bacteriology* 195.4 (2012): 726-32. Web.
2. Liu, Huanting, Kerry Woznica, Gemma Catton, Amanda Crawford, Nigel Botting, and James H. Naismith. "Structural and Kinetic Characterization of Quinolinate Phosphoribosyltransferase (hQPRTase) from Homo Sapiens." *Journal of Molecular Biology* 373.3 (2007): 755-63. Web.
3. Finn, Robert D., Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Jaina Mistry, Alex L. Mitchell, Simon C. Potter, Marco Punta, Matloob Qureshi, Amaia Sangrador-Vegas, Gustavo A. Salazar, John Tate, and Alex Bateman. "The Pfam Protein Families Database: Towards a More Sustainable Future." *Nucleic Acids Res Nucleic Acids Research* 44.D1 (2015): n. pag. Web.
4. Ollis, David L., Eong Cheah, Miroslaw Cygler, Bauke Dijkstra, Felix Frolow, Sybille M. Franken, Michal Harel, S. Jamse Remington, Israel Silman, Joseph Schrag, Joel L. Sussman, Koen H.g. Verschueren, and Adrian Goldman. "The α / β Hydrolase Fold." *"Protein Engineering, Design and Selection" Protein Eng Des Sel* 5.3 (1992): 197-211. Web
5. Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, Z. Zheng, and S. H. Bryant. "CDD: A Conserved Domain Database for Protein Classification." *Nucleic Acids Research* 43.Database Issue (2015): 222-26. Web.
6. Wierenga, R.k. "The TIM-barrel Fold: A Versatile Framework for Efficient Enzymes." *FEBS Letters* 492.3 (2001): 193-98. Web.
7. Cao, Hong, Beth L. Pietrak, and Charles Grubmeyer. "Quinolinate Phosphoribosyltransferase: Kinetic Mechanism for a Type II PRTase †." *Biochemistry* 41.10 (2002): 3520-528. Web.
8. Majorek, Karolina A., Misty L. Kuhn, Maksymilian Chruszcz, Wayne F. Anderson, and Wladek Minor. "Double Trouble-Buffer Selection and His-tag Presence May Be Responsible for Nonreproducibility of Biomedical Experiments." *Protein Science* 23.10 (2014): 1359-368. Web.
9. Qazi, S. Junaid S., Raymond Chew, Denice C. Bay, and Raymond J. Turner. "Structural and Functional Comparison of Hexahistidine Tagged and Untagged Forms of Small Multidrug Resistance Protein, EmrE." *Biochemistry and Biophysics Reports* 1 (2015): 22-32. Web
10. DeSantis, Kara, Aaron Reed, Raneen Rahhal, and Jeff Reinking. "Use of Differential Scanning Fluorimetry as a High-throughput Assay to Identify Nuclear Receptor Ligands." *Nuclear Receptor Signaling Nucl Recept Signal* 10 (2012): 1-5. Web.
11. Dong, Aiping, Xiaohui Xu, Aled M. Edwards, Changsoo Chang, Maksymilian Chruszcz, Marianne Cuff, Marcin Cymborowski, Rosa Di Leo, Olga Egorova, Elena Evdokimova, Ekaterina Filippova, Jun Gu, Jennifer Guthrie, Alexandr Ignatchenko, Andrzej Joachimiak, Natalie Klostermann, Youngchang Kim, Yuri Korniyenko, Wladek Minor, Qiuni Que, Alexei Savchenko, Tatiana Skarina, Kemin Tan, Alexander

Yakunin, Adelinda Yee, Veronica Yim, Rongguang Zhang, Hong Zheng, Masato Akutsu, Cheryl Arrowsmith, George V. Avvakumov, Alexey Bochkarev, Lars-Göran Dahlgren, Sirano Dhe-Paganon, Slav Dimov, Ludmila Dombrovski, Patrick Finerty, Susanne Flodin, Alex Flores, Susanne Gräslund, Martin Hammerström, Maria Dolores Herman, Bum-Soo Hong, Raymond Hui, Ida Johansson, Yongson Liu, Martina Nilsson, Lyudmila Nedyalkova, Pär Nordlund, Tomas Nyman, Jinrong Min, Hui Ouyang, Hee-Won Park, Chao Qi, Wael Rabeh, Limin Shen, Yang Shen, Deepthi Sukumard, Wolfram Tempel, Yufeng Tong, Lionel Tresagues, Masoud Vedadi, John R. Walker, Johan Weigelt, Martin Welin, Hong Wu, Ting Xiao, Hong Zeng, and Haizhong Zhu. "In Situ Proteolysis for Protein Crystallization and Structure Determination." *Nature Methods Nat Meth* 4.12 (2007): 1019-021. Web

12. Mao, F., P. Dam, J. Chou, V. Olman, and Y. Xu. "DOOR: A Database for Prokaryotic Operons." *Nucleic Acids Research* 37.Database (2009): n. pag. Web.
13. Krissinel, Evgeny, and Kim Henrick. "Inference of Macromolecular Assemblies from Crystalline State." *Journal of Molecular Biology* 372.3 (2007): 774-97. Web.
14. Sharma, Vivek, Charles Grubmeyer, and James C. Sacchettini. "Crystal Structure of Quinolinic Acid Phosphoribosyltransferase from Mycobacterium Tuberculosis: A Potential TB Drug Target." *Structure* 6.12 (1998): 1587-599. Web.
15. Magrane, M., and U. Consortium. "UniProt Knowledgebase: A Hub of Integrated Protein Data." *Database* 2011.0 (2011): n. pag. Web
16. Altschul, Stephen F., John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schaffer, and Yi-Kuo Yu. "Protein Database Searches Using Compositionally Adjusted Substitution Matrices." *FEBS Journal* 272.20 (2005): 5101-109. Web.
17. Mandel, M., and A. Higa. "Calcium-dependent Bacteriophage DNA Infection." *Journal of Molecular Biology* 53.1 (1970): 159-62. Web
18. Die, I. M. Van, H. E. N. Bergmans, and W. P. M. Hoekstra. "Transformation In Escherichia Coli: Studies On The Role Of The Heat Shock In Induction Of Competence." *Microbiology* 129.3 (1983): 663-70. Web.
19. Niesen, Frank H., Helena Berglund, and Masoud Vedadi. "The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability." *Nat Protoc Nature Protocols* 2.9 (2007): 2212-221. Web.
20. Berggren, Kiera, Thomas H. Steinberg, Wendy M. Lauber, James A. Carroll, Mary F. Lopez, Elena Chernokalskaya, Lynn Zieske, Zhenjun Diwu, Richard P. Haugland, and Wayne F. Patton. "A Luminescent Ruthenium Complex for Ultrasensitive Detection of Proteins Immobilized on Membrane Supports." *Analytical Biochemistry* 276.2 (1999): 129-43. Web.
21. Minor, Wladek, Marcin Cymborowski, Zbyszek Otwinowski, and Maksymilian Chruszcz. "HKL -3000: The Integration of Data Reduction and Structure Solution – from Diffraction Images to an Initial Model in Minutes." *Acta Crystallogr D Biol Cryst Acta Cryst D Acta Crystallogr Sect D Biol Crystallogr Acta Crystallogr Sect D Acta Crystallogr D Biol Crystallogr Acta Crystallographica Section D Biological Crystallography Acta Cryst Sect D Acta Crystallogr D* 62.8 (2006): 859-66. Web.
22. Vagin, A., and A. Teplyakov. "MOLREP: An Automated Program for Molecular Replacement." *J Appl Cryst Journal of Applied Crystallography J Appl Crystallogr* 30.6 (1997): 1022-025. Web.

23. Murshudov, Garib N., Alexei A. Vagin, Andrey Lebedev, Keith S. Wilson, and Eleanor J. Dodson. "Efficient Anisotropic Refinement of Macromolecular Structures Using FFT." *Acta Crystallogr D Biol Cryst Acta Cryst D Acta Crystallogr Sect D Biol Crystallogr Acta Crystallogr Sect D Acta Crystallogr D Biol Crystallogr Acta Crystallographica Section D Biological Crystallography Acta Cryst Sect D Acta Crystallogr D* 55.1 (1999): 247-55. Web.
24. Emsley, Paul, and Kevin Cowtan. "Coot : Model-building Tools for Molecular Graphics." *Acta Crystallogr D Biol Cryst Acta Cryst D Acta Crystallogr Sect D Biol Crystallogr Acta Crystallogr Sect D Acta Crystallogr D Biol Crystallogr Acta Crystallographica Section D Biological Crystallography Acta Cryst Sect D Acta Crystallogr D* 60.12 (2004): 2126-132. Web.
25. Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. "MolProbity : All-atom Structure Validation for Macromolecular Crystallography." *International Tables Online International Tables for Crystallography Crystallography of Biological Macromolecules* (2012): 694-701. Web.
26. Zheng, Heping, Mahendra D. Chordia, David R. Cooper, Maksymilian Chruszcz, Peter Müller, George M. Sheldrick, and Wladek Minor. "Validation of Metal-binding Sites in Macromolecular Structures with the CheckMyMetal Web Server." *Nat Protoc Nature Protocols* 9.1 (2013): 156-70. Web.
27. Holm, L., and P. Rosenstrom. "Dali Server: Conservation Mapping in 3D." *Nucleic Acids Research* 38.Web Server (2010): n. pag. Web.
28. Slabinski, L., L. Jaroszewski, L. Rychlewski, I. A. Wilson, S. A. Lesley, and A. Godzik. "XtalPred: A Web Server for Prediction of Protein Crystallizability." *Bioinformatics* 23.24 (2007): 3403-405. Web.
29. Gasteiger, Elisabeth, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. "Protein Identification and Analysis Tools on the ExPASy Server." *The Proteomics Protocols Handbook* (2005): 571-607. Web.
30. Kelley, Lawrence A., Stefans Mezulis, Christopher M. Yates, Mark N. Wass, and Michael J E Sternberg. "The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis." *Nat Protoc Nature Protocols* 10.6 (2015): 845-58. Web.
31. Bordoli, Lorenza, Florian Kiefer, Konstantin Arnold, Pascal Benkert, James Battey, and Torsten Schwede. "Protein Structure Homology Modeling Using SWISS-MODEL Workspace." *Nat Protoc Nature Protocols* 4.1 (2008): 1-13. Web.
32. Delano, Warren, and Schrödinger, LLC. *The PyMOL Molecular Graphics System*. Computer software. *PyMOL*. Vers. 1.3. Schrödinger, LLC, 2011. Web. 2 Jan. 2016.
33. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. Mcwilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. "Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology* 7.1 (2014): 539. Web.
34. Robert, X., and P. Gouet. "Deciphering Key Features in Protein Structures with the New ENDscript Server." *Nucleic Acids Research* 42.W1 (2014): n. pag. Web.
35. Ashkenazy, H., E. Erez, E. Martz, T. Pupko, and N. Ben-Tal. "ConSurf 2010: Calculating Evolutionary Conservation in Sequence and Structure of Proteins and Nucleic Acids." *Nucleic Acids Research* 38.Web Server (2010): n. pag. Web.

36. Cook, Paul F., and W. W. Cleland. *Enzyme Kinetics and Mechanism*. London: Garland Science, 2007. Print.

CHAPTER 3

STRUCTURAL STUDIES OF NICOTINATE MONONUCLEOTIDE ADENYLYLTRANSFERASE (NADD) OF THE GAS QUINOLINATE-SALVAGE PATHWAY

3.1 INTRODUCTION

SpNadD is the second of the trio of proteins, within the QSP, that is responsible for the conversion of NaMN and ATP into NaAD (1) (Figure 3.1). According to the Database for Essential Genes (DEGG) and available literature, the expression of the *nadD* gene is integral for the survival of various pathogenic bacteria (1, 2). Moreover, the fact that human homolog of NadD and spNadD have only 20% sequence identity makes the *S. pyogenes* protein an attractive target for development of antimicrobial compounds. NadD is a member of the nucleotidyl transferase superfamily of proteins (3). A signature trademark of this superfamily is the presence of a (T/H)XGH motif that has been observed, to play a part in the binding of ATP within the active site (Figure 3.2) (4). Other members of this superfamily include adenyltransferases, tRNA synthetases, panthothenate synthetase, ATP sulfurylases and cytidylyltransferase. As reported by Interpro, a common characteristic feature of this superfamily is the presence of the Rossman-like $\alpha/\beta/\alpha$ sandwich fold (5). Despite the fact that NadD proteins have very similar overall fold they may form different oligomeric structures. For example these proteins were reported to have dimeric *Staphylococcus aureus* NadD (PDB code 2H29), tetrameric *Bacillus subtilis* NadD (PDB code: 1KAM) or hexameric *Homo sapiens* NadD (PDB code: 1KR2) quaternary

structures (4, 6, 7). The *Escherichia coli* homolog of NadD (ecNadD; PDB code: 1K4M), has been observed to come in two forms: monomeric in the apo, and trimeric in the substrate-bound form (9). However, the trimeric form has only been observed in crystal (9) (Figure 3.3).

This structure of spNadD has yet to be determined. This chapter will present results of protein production, crystallization, thermal stability, initial structural results, and homology modeling.

3.2 Results and Discussion

Protein Production

Recombinant spNadD was successfully expressed and isolated at a yield of ~50 mg/ L of culture. SDS-PAGE results indicated that isolation of spNadD was approximately 95% pure after gel filtration. Polyhisitdine tag cleavage experiments were successful for spNadD, and it was the only QSP for which the purification tag could be removed. Results of gel filtration experiments suggests that spNadC form a trimer in solution. These results were consistent with the dynamic light scattering and native gel electrophoresis experiments. The molecular weight of spNadD, as determined by DLS, was approximately 72 (± 9.7) kDa, while the molecular weight of a single spNadD chain is 24 kDa. Native Gels tested to determine QSE complex formation also indicate that spNadD is trimer and does not form a complex with either of the other QSEs (See Chapter 2, Figure 2.3).

THERMAL STABILITY

Analysis of the protein thermal stability in solutions with varying pH and NaCl concentration showed that spNadD was most stable in conditions with high salinity and alkaline pH (Figure 3.4). There was no significant difference in the thermal stability of spNadD with the purification tag compared to the protein with the cleaved tag (Figure 3.5).

The thermal stability studies revealed that various compounds may increase the melting temperature of spNadD. For example, alanine, aminobiphenyl (in 15% DMSO), cysteine, ferrous sulfate, malate, sodium pyruvate, and succinate promoted NadD stability up to 60 °C (relative to the control conditions in which protein melted at 50 °C). Cadmium chloride, Fast violet B salt, magnesium chloride, and sodium bromide promoted stability up to 62 °C. Potassium tetrachloroplatinate, provided the greatest amount of stability at 64 °C (Figure 3.6).

CRYSTALLIZATION

Crystals of spNadD with and without purification tag may be obtained using the following conditions: 1) 0.1 M succinic acid, 15% w/v PEG 3350 2) 20% w/v PEG 3350 and 3) 20% w/v PEG 3350, 0.2 M $MgCl_2$, 0.1 M Tris (pH 8.5) (Figure 3.7). Crystals for this protein, in all conditions, have long crystallization periods ranging from 2 weeks to 4 months. Crystals in the succinate condition produced irregularly-shaped clusters. Crystals in the conditions in 2 and 3 produced long, hexagonal prism-shaped crystals. Crystals obtained using condition #3 produced the best diffraction data. Various attempts to improve quality of the crystals failed. For example, attempts to crystallize protein in presence of substrate and substrate analogs were unsuccessful. Addition of the compounds

that were identified to increase the thermal stability of NadD also did not improve the quality of crystals. The protein was successfully modified using the reductive methylation of lysine. However, the methylated protein could not be crystallized. Work is currently being conducted to further optimize the conditions for crystal growth and identify solutions that would promote crystal growth when the protein is in complex with its substrates.

ANALYSIS OF NADD SEQUENCE

According to the DOOR[2] the *nadD* gene is located further downstream from *nadC* gene, with respect to the origin, at position 275946-276578. SpNadD is the smallest of the QSPs with only 210 residues in the polypeptide chain. Sequence analysis results suggest that the areas with the highest sequence conservation, between NadD homologs with known structures, include residues: 30-44, 60, 78-81, 103-110, 128, 131, 138, 144, 154, 155, 182, and 199, 202, 208, and 209. Most of the conserved residues between 30- 138 are affiliated with the active site. According to Uniprot and Pfam spNadD contains the CTP transferase-like domain that is composed of residues 28-183. NCBI Protein BLAST results provide predicted locations of two major segments within the sequence: the active site (substrate stabilizing) segment, and the (T/H)XGH motif. The sequence analysis suggests that the active site residues include: L29, **G30**, **G31**, N32, **H37**-**H40** (T/H) XGH motif), **V43**, P64, S105, **Y106**, **T107**, F125, I127, **G128**, **D130**, **M131**, D138, **W138**, **H139**, Q154, **R155**, and I176. The (T/H)XGH motif residues in spNadD include **H37**, N38, A39, and **H40**. The residues are highlighted in bold to indicate that they are highly conserved. Refer to Figure 3.8 for spNadD sequence alignment and residue conservation results with NadD homologs with known structures.

Initial Structural Analysis

Crystal Structure

The best crystal of spNadD with cleaved purification tag diffracted only to 8.5 Å resolution. The low resolution data suggests that the protein crystallized in a monoclinic system and $P2_1$ space group. The unit cell parameters were identified as: a =104Å, b =103 Å, c =114 Å, β =108º. Analysis of the probable solvent content using the Matthews coefficient indicates that between 4 to 12 copies of spNadD are present in the asymmetric unit (10). Gel filtration analysis, native gel analysis, and DSL results indicate that spNadD is a trimer. Therefore, taking into account information on the crystal symmetry, the oligomeric state of the protein and the solvent content most likely the unit cell contains 6, 9 or 12 protein chains in the asymmetric unit.

Homology Modeling and Analysis of the Model

The PHYRE2 server was used to calculate homology models for the spNadD. PHYRE2 reports the model results with structural quality parameters for confidence, coverage, and identity. PHYRE2 defines confidence as "the probability that your (input) sequence and the template are homologous". The top four homology models used as templates were from *Bacillus anthracis* (baNadD; PDB code: 3E27), *Bacillus subtilis* (bsNadD; PDB code: 1KAM), *Staphylococcus aureus* (saNadD; PDB code: 2H29), and *Pseudomonas aeruginosa* (paNadD; PDB: 1YUN). Each of the homologs had confidence values of 100%, coverage of ≥ 88%, and identity of ≥ 33%. Cartoon representations of these structures are shown in Figures 3.9 and 3.10. RMSD values between the NadD homology model, built from a *Bacillus anthracis* template (PDB code: 3E27), against the

top four homology models were: 1.2 Å for the saNadD model, 1.5 Å for the ecNadD model, 1.5 Å for the bsNadD model and 1.8Å for the paNadD model. Structural comparisons show the greatest variability in the orientation of the loop regions, but the the secondary structure elements forming structure the core of the protein are the same in all generated models.

NadD Active Site Prediction

To gain insight into the active site location we compared structures of spNadD homologs (saNadD, baNadD, mtNadD and bsNadD). It was observed that each monomer, shows NaAD bound in similar locations. The adenylate moiety of NaAD is observed to bind closer to the N-terminus which allows for stabilization by the highly conserved (T/H)XGH motif, described above. For the spNadD model and the homologs the motif is commonly found with in the first 20 residues, from the N-terminus of the structure. Specifically, the motif for spNadD is "HNAH", "HTAH" for saNadD, "HNGH" for bsNadD, "HYAH" for saNadD, and for baNadD and mtNadD for it is "HGYH". The notable aspect of this motif is the orientation of both of the histidine residues, as they are always oriented toward the adenlylate, located deep within the site. The acidic residues of the motif are oriented toward the protein surface. Refer to Figure 3.11 and 3.12 for a cartoon depiction of the binding pocket in open and closed conformations. Superposition results for spNadD homology model versus the structural homologs reveal that there is residue consistency, compared to spNadD residues described in Section 3.2.4, between structures with variability in residue orientation dependent upon what is in the active site. The orientation of the baNadD structure (PDB code: 3HFJ) with the inhibitor-bound does show a major difference in residue orientation, but the binding site for the inhibitor molecule is in a different location than the NaAD site. Contrary to this exception, the

remaining structure results give confidence that the homology model could be an accurate depiction of this protein and it is assumed that spNadD will contain the same active site location.

Conclusions and Future Directions

NadD continues to be the most elusive of the QSPs for structure determination, yet may be one of the most critical primary targets for inhibitor development (8). Experimental data has determined is that the protein is extremely stable, but difficult to crystalize efficiently. Similar to results obtained by Zhang *et al.*, from a homologous NadD construct from *E. coli*, show that it can take weeks to grow crystals (9). Difficulties with inconsistency, with manual preparation, can make it a challenge to quickly observe changes to crystal formation, and optimize in an efficient manner. To overcome this two strategies were employed to decrease general entropy of the protein in order to improve the likelihood of crystallization. In first strategy, we decided methylate the lysine residues in all of the QSPs in order to improve the probability of crystallization and/or change crystal morphology in order to improve diffraction for higher resolution results. The second strategy targeted the mutation of the lysine residues to improve the formation of crystal contacts. We utilized the SERp server (UCLA) to identify target lysine residues for mutation in order to decrease overall protein entropy (11). Results from the server indicate that lysine residues 44-45 and residues 210-211 (in DNA 2.0 optimized sequence) were prime targets for mutation (11). Using the NCBI Protien BLAST server, these residues were also determined to not interfere with residues involved with active site binding (11, 12, 13). Future experiments will utilize molecular cloning to mutate lysine residues into alanine residues thus creating three NadD mutant constructs. These NadD constructs will

have mutations at residues 44-45 only, 210-211 only, and a double mutant will have substitutions at both sites at both sites. By decreasing entropy we hope to create protein constructs that will afford a more efficient means of crystal formation, for higher quality crystals for eventual x-ray analysis.

3.3 Materials and Methods

Protein Production and Crystallization

All protocols for protein production, isolation, and crystallization experiments are consistent with the experimental protocol outlined previously in Chapter 2, pages 27-30. In addition, for the purpose of crystallization spNadD was methylated. The reductive methylation of lysine was performed using a protocol outline previously (14). Briefly, each the QSPs were stored at -80 °C in 2 mg/mL concentrations for crystallization and were thawed, on ice, prior to the methylation experiment. The protein solutions were initially diluted two-fold in dialysis buffer (50 mM Tris (pH 7.5), 150 mM NaCl, and 5 mM 2-mercapto-ethanol). The solution was then placed in an ice bath for 30 min. After ice bath incubation, 50 µL of 85 mM formaldehyde (diluted from 1 M stock in dialysis buffer) was added to the protein solution and was gently mixed. Afterward 25 µL of 1 M Borane-Dimethylamine (BDA) complex was added to the solution, mixed gently and returned to the ice bath for two hours. After the two hour incubation, the formaldehyde and BDA was added again to the protein solution at the volumes described above, and was returned to the ice bath for an additional hour. Upon completion of the hour long incubation 20 µL more of BDA was added and left at 4 °C for 12-16 hours. After the overnight incubation, 1-2 mg of solid glycine was added and was placed in the ice bath for an additional hour. Upon

completion the solution was concentrated and buffer exchanged in preparation for MALDI analysis for confirmation of experiment success.

Dynamic Light Scattering

All protocols for dynamic light scattering experiments are consistent with the protocols outlined previously in Chapter 2, page 30.

Differential Scanning Fluorimetry

SpNadD thermal stability experiments were conducted using the Differential Scanning Fluorimetry strategy outlined in Chapter 1, pages 30-31.

The thermal stability studies of spNadD in the presence of various small molecular compounds were performed using the same approach as in the case of the pH vs. salinity studies. The difference in the protocols was that the pH and salinity buffer solution was exchanged with 10 µL of buffered small molecule solutions (0.05-0.1 M Tris pH 7.5). Contents of each of the 384 conditions are contained in **Appendix I**. This protocol was prepared by Dr. Nicholas Mank, which was adapted from the original work of Niesen *et al.* (15). The conditions were prepared by an undergraduate volunteer Vincent Klapper.

Data collection and processing

SpNadD crystals were cryo-cooled in 10 % ethylene glycol, frozen in liquid nitrogen and placed in a puck. All spNadD data were collected at the Advanced Photon Source at Argonne National Laboratory in Lemont, IL. The diffraction experiments were conducted remotely on the 22ID beamline run by the Southeast Regional Collaborative

Access Team. All diffraction data was collected at 100 K, and results were processed using HKL-3000 (16).

Other Computational Methods

All other computational methods used are consistent with the methods used in Chapter 2, page 33.

3.4 TABLES AND FIGURES

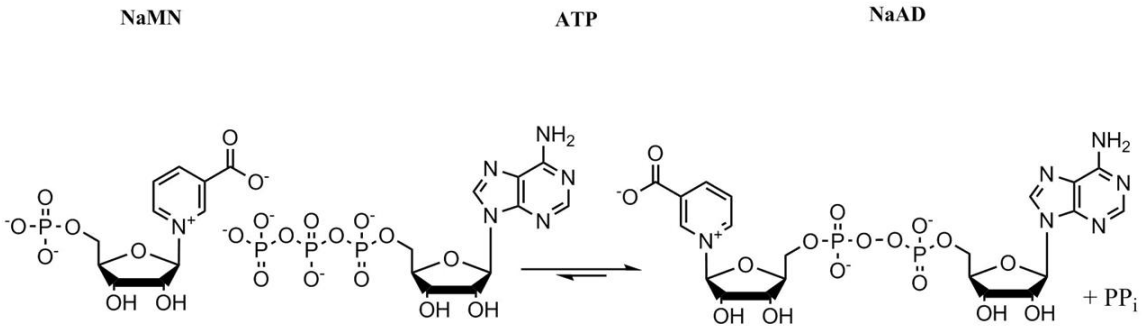


Figure 3.1: General reaction for spNadD: This figure is a schematic of the reaction catalyzed by NadD.

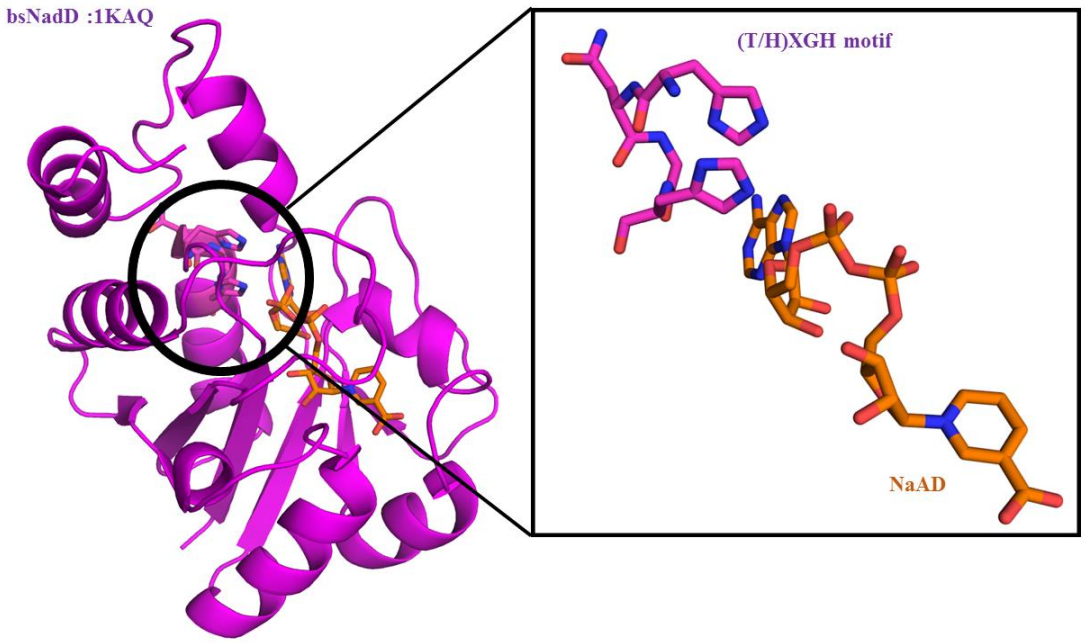


Figure 3.2: The (T/H)XGH motif. Cartoon representation of a single *Bacillus subtilis* NadD chain. The magnified region depicts the position of the (T/H)XGH motif and its orientation to the adenylate moiety of NaAD.

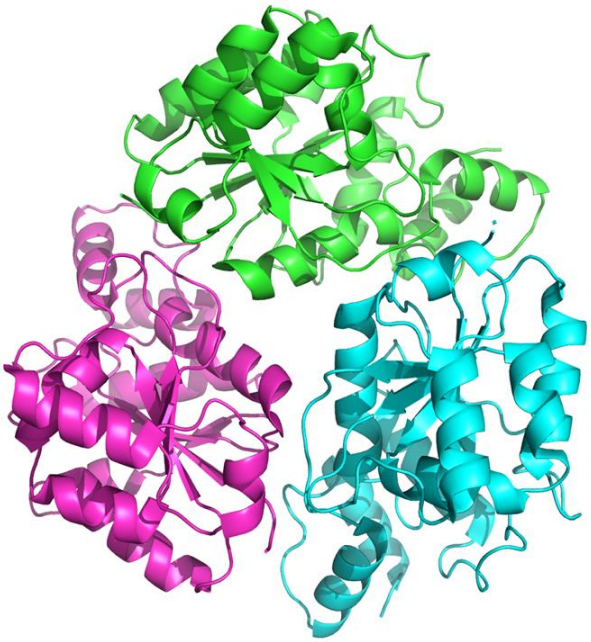


Figure 3.3: Proposed structure of spNadD. Trimer structure from ecNadD (PDB code: 1K4M, 37 % identity) is suggested to be the predicted oligomeric assembly for spNadD.

< 30 °C 30's 40's 50 °C 51 °C 52 °C 53 °C 54°C 55 °C 56 °C 57 °C 58 °C 59 °C

Uncut sp**NadD** (2 mg/mL stock)

	1	2	3	4	5	6	7	8	9	10	11	12
A	Acetate pH 4.0 0 mM NaCl	Acetate pH 4.5, 0 mM NaCl	Acetate pH 5.0, 0 mM NaCl	Bis-Tris pH 5.5, 0 mM NaCl	Bis-Tris pH 6.0, 0 mM NaCl	Bis-Tris pH 6.5, 0 mM NaCl	Tris pH 7.0, 0 mM NaCl	Tris pH 7.5, 0 mM NaCl	Tris pH 8.0, 0 mM NaCl	CHES pH 8.5, 0 mM NaCl	CHES pH 9.0, 0 mM NaCl	CHES pH 9.5, 0 mM NaCl
B	Acetate pH 4.0, 50 mM NaCl	Acetate pH 4.5, 50 mM NaCl	Acetate pH 5.0, 50 mM NaCl	Bis-Tris pH 5.5, 50 mM NaCl	Bis-Tris pH 6.0, 50 mM NaCl	Bis-Tris pH 6.5, 50 mM NaCl	Tris pH 7.0, 50 mM NaCl	Tris pH 7.5, 50 mM NaCl	Tris pH 8.0, 50 mM NaCl	CHES pH 8.5, 50 mM NaCl	CHES pH 9.0, 50 mM NaCl	CHES pH 9.5, 50 mM NaCl
C	Acetate pH 4.0, 100 mM NaCl	Acetate pH 4.5, 100 mM NaCl	Acetate pH 5.0, 100 mM NaCl	Bis-Tris pH 5.5, 100 mM NaCl	Bis-Tris pH 6.0, 100 mM NaCl	Bis-Tris pH 6.5, 100 mM NaCl	Tris pH 7.0, 100 mM NaCl	Tris pH 7.5, 100 mM NaCl	Tris pH 8.0, 100 mM NaCl	CHES pH 8.5, 100 mM NaCl	CHES pH 9.0, 100 mM NaCl	CHES pH 9.5, 100 mM NaCl
D	Acetate pH 4.0, 150 mM NaCl	Acetate pH 4.5, 150 mM NaCl	Acetate pH 5.0, 150 mM NaCl	Bis-Tris pH 5.5, 150 mM NaCl	Bis-Tris pH 6.0, 150 mM NaCl	Bis-Tris pH 6.5, 150 mM NaCl	Tris pH 7.0, 150 mM NaCl	Tris pH 7.5, 150 mM NaCl	Tris pH 8.0, 150 mM NaCl	CHES pH 8.5, 150 mM NaCl	CHES pH 9.0, 150 mM NaCl	CHES pH 9.5, 150 mM NaCl
E	Acetate pH 4.0, 200 mM NaCl	Acetate pH 4.5, 200 mM NaCl	Acetate pH 5.0, 200 mM NaCl	Bis-Tris pH 5.5, 200 mM NaCl	Bis-Tris pH 6.0, 200 mM NaCl	Bis-Tris pH 6.5, 200 mM NaCl	Tris pH 7.0, 200 mM NaCl	Tris pH 7.5, 200 mM NaCl	Tris pH 8.0, 200 mM NaCl	CHES pH 8.5, 200 mM NaCl	CHES pH 9.0, 200 mM NaCl	CHES pH 9.5, 200 mM NaCl
F	Acetate pH 4.0, 250 mM NaCl	Acetate pH 4.5, 250 mM NaCl	Acetate pH 5.0, 250 mM NaCl	Bis-Tris pH 5.5, 250 mM NaCl	Bis-Tris pH 6.0, 250 mM NaCl	Bis-Tris pH 6.5, 250 mM NaCl	Tris pH 7.0, 250 mM NaCl	Tris pH 7.5, 250 mM NaCl	Tris pH 8.0, 250 mM NaCl	CHES pH 8.5, 250 mM NaCl	CHES pH 9.0, 250 mM NaCl	CHES pH 9.5, 250 mM NaCl
G	Acetate pH 4.0, 500 mM NaCl	Acetate pH 4.5, 500 mM NaCl	Acetate pH 5.0, 500 mM NaCl	Bis-Tris pH 5.5, 500 mM NaCl	Bis-Tris pH 6.0, 500 mM NaCl	Bis-Tris pH 6.5, 500 mM NaCl	Tris pH 7.0, 500 mM NaCl	Tris pH 7.5, 500 mM NaCl	Tris pH 8.0, 500 mM NaCl	CHES pH 8.5, 500 mM NaCl	CHES pH 9.0, 500 mM NaCl	CHES pH 9.5, 500 mM NaCl
H	Acetate pH 4.0, 1000 mM NaCl	Acetate pH 4.5, 1000 mM NaCl	Acetate pH 5.0, 1000 mM NaCl	Bis-Tris pH 5.5, 1000 mM NaCl	Bis-Tris pH 6.0, 1000 mM NaCl	Bis-Tris pH 6.5, 1000 mM NaCl	Tris pH 7.0, 1000 mM NaCl	Tris pH 7.5, 1000 mM NaCl	Tris pH 8.0, 1000 mM NaCl	CHES pH 8.5, 1000 mM NaCl	CHES pH 9.0, 1000 mM NaCl	CHES pH 9.5, 1000 mM NaCl

Figure 3.4: Thermal stability results for uncut spNadD. This figure shows the results of average denaturation temperatures, for his-tagged spNadD, in various pH and salinity conditions during thermal stability experiments. Each DSF experiment was done in triplicate.

< 30 °C 30's 40's 50 °C 51 °C 52 °C 53 °C 54°C 55 °C 56 °C 57 °C 58 °C 59 °C

Cut sp**NadD** (2 mg/mL stock)

	1	2	3	4	5	6	7	8	9	10	11	12
A	Acetate pH 4.0 0 mM NaCl	Acetate pH 4.5, 0 mM NaCl	Acetate pH 5.0, 0 mM NaCl	Bis-Tris pH 5.5, 0 mM NaCl	Bis-Tris pH 6.0, 0 mM NaCl	Bis-Tris pH 6.5, 0 mM NaCl	Tris pH 7.0, 0 mM NaCl	Tris pH 7.5, 0 mM NaCl	Tris pH 8.0, 0 mM NaCl	CHES pH 8.5, 0 mM NaCl	CHES pH 9.0, 0 mM NaCl	CHES pH 9.5, 0 mM NaCl
B	Acetate pH 4.0, 50 mM NaCl	Acetate pH 4.5, 50 mM NaCl	Acetate pH 5.0, 50 mM NaCl	Bis-Tris pH 5.5, 50 mM NaCl	Bis-Tris pH 6.0, 50 mM NaCl	Bis-Tris pH 6.5, 50 mM NaCl	Tris pH 7.0, 50 mM NaCl	Tris pH 7.5, 50 mM NaCl	Tris pH 8.0, 50 mM NaCl	CHES pH 8.5, 50 mM NaCl	CHES pH 9.0, 50 mM NaCl	CHES pH 9.5, 50 mM NaCl
C	Acetate pH 4.0, 100 mM NaCl	Acetate pH 4.5, 100 mM NaCl	Acetate pH 5.0, 100 mM NaCl	Bis-Tris pH 5.5, 100 mM NaCl	Bis-Tris pH 6.0, 100 mM NaCl	Bis-Tris pH 6.5, 100 mM NaCl	Tris pH 7.0, 100 mM NaCl	Tris pH 7.5, 100 mM NaCl	Tris pH 8.0, 100 mM NaCl	CHES pH 8.5, 100 mM NaCl	CHES pH 9.0, 100 mM NaCl	CHES pH 9.5, 100 mM NaCl
D	Acetate pH 4.0, 150 mM NaCl	Acetate pH 4.5, 150 mM NaCl	Acetate pH 5.0, 150 mM NaCl	Bis-Tris pH 5.5, 150 mM NaCl	Bis-Tris pH 6.0, 150 mM NaCl	Bis-Tris pH 6.5, 150 mM NaCl	Tris pH 7.0, 150 mM NaCl	Tris pH 7.5, 150 mM NaCl	Tris pH 8.0, 150 mM NaCl	CHES pH 8.5, 150 mM NaCl	CHES pH 9.0, 150 mM NaCl	CHES pH 9.5, 150 mM NaCl
E	Acetate pH 4.0, 200 mM NaCl	Acetate pH 4.5, 200 mM NaCl	Acetate pH 5.0, 200 mM NaCl	Bis-Tris pH 5.5, 200 mM NaCl	Bis-Tris pH 6.0, 200 mM NaCl	Bis-Tris pH 6.5, 200 mM NaCl	Tris pH 7.0, 200 mM NaCl	Tris pH 7.5, 200 mM NaCl	Tris pH 8.0, 200 mM NaCl	CHES pH 8.5, 200 mM NaCl	CHES pH 9.0, 200 mM NaCl	CHES pH 9.5, 200 mM NaCl
F	Acetate pH 4.0, 250 mM NaCl	Acetate pH 4.5, 250 mM NaCl	Acetate pH 5.0, 250 mM NaCl	Bis-Tris pH 5.5, 250 mM NaCl	Bis-Tris pH 6.0, 250 mM NaCl	Bis-Tris pH 6.5, 250 mM NaCl	Tris pH 7.0, 250 mM NaCl	Tris pH 7.5, 250 mM NaCl	Tris pH 8.0, 250 mM NaCl	CHES pH 8.5, 250 mM NaCl	CHES pH 9.0, 250 mM NaCl	CHES pH 9.5, 250 mM NaCl
G	Acetate pH 4.0, 500 mM NaCl	Acetate pH 4.5, 500 mM NaCl	Acetate pH 5.0, 500 mM NaCl	Bis-Tris pH 5.5, 500 mM NaCl	Bis-Tris pH 6.0, 500 mM NaCl	Bis-Tris pH 6.5, 500 mM NaCl	Tris pH 7.0, 500 mM NaCl	Tris pH 7.5, 500 mM NaCl	Tris pH 8.0, 500 mM NaCl	CHES pH 8.5, 500 mM NaCl	CHES pH 9.0, 500 mM NaCl	CHES pH 9.5, 500 mM NaCl
H	Acetate pH 4.0, 1000 mM NaCl	Acetate pH 4.5, 1000 mM NaCl	Acetate pH 5.0, 1000 mM NaCl	Bis-Tris pH 5.5, 1000 mM NaCl	Bis-Tris pH 6.0, 1000 mM NaCl	Bis-Tris pH 6.5, 1000 mM NaCl	Tris pH 7.0, 1000 mM NaCl	Tris pH 7.5, 1000 mM NaCl	Tris pH 8.0, 1000 mM NaCl	CHES pH 8.5, 1000 mM NaCl	CHES pH 9.0, 1000 mM NaCl	CHES pH 9.5, 1000 mM NaCl

Figure 3.5: Thermal shift results for cut spNadD. This figure shows the results of average denaturation temperatures, for spNadD with a cleaved his-tag, in various pH and salinity conditions during thermal stability experiments. Each DSF experiment was done in triplicate.

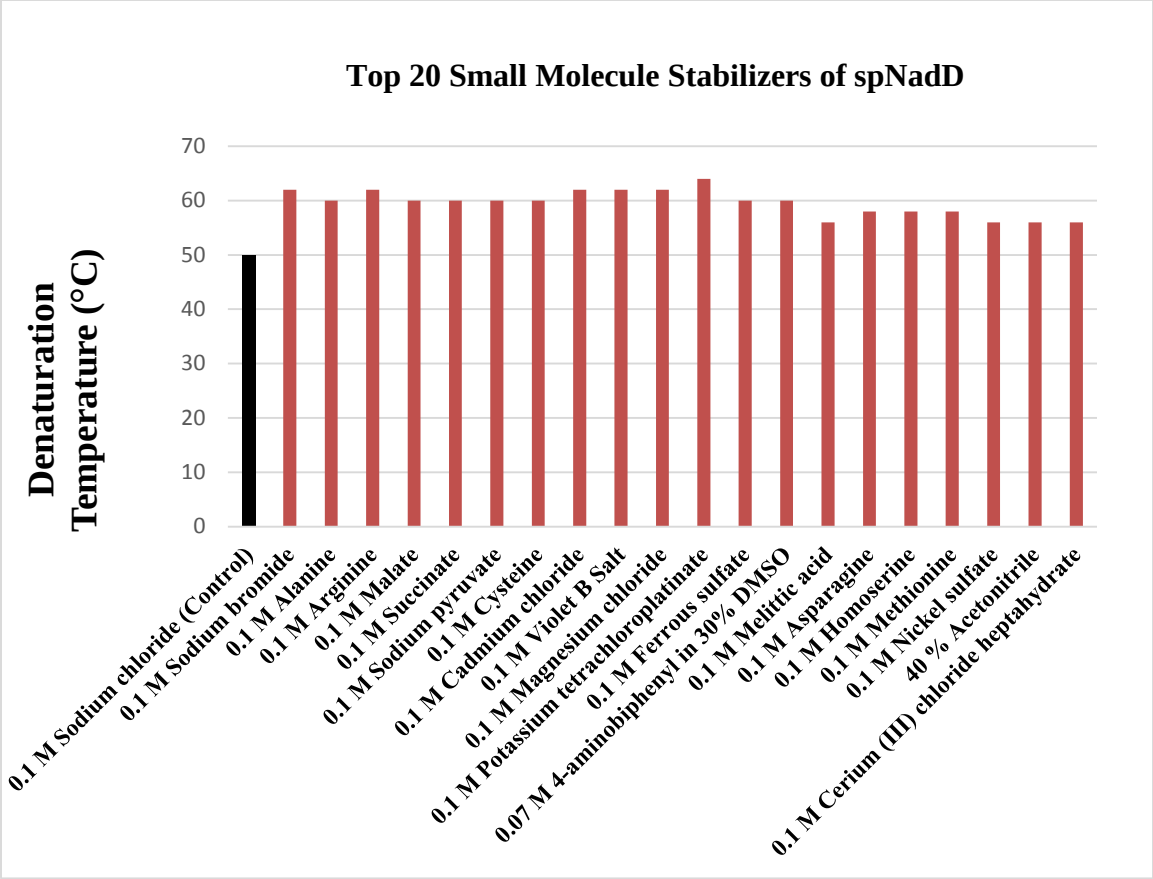


Figure 3.6: NadD small molecule stability. Thermal stability results of the top 20 stabilizers out of 396 total small molecule conditions.

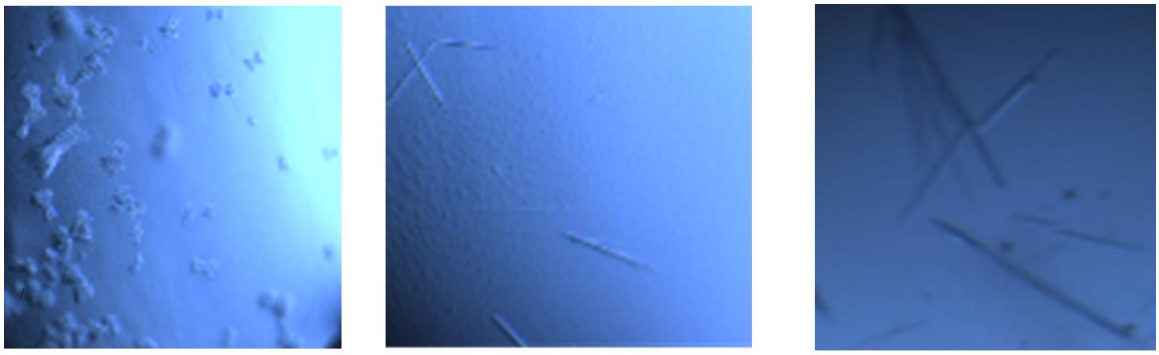


Figure 3.7. NadD crystallization results. These are images of NadD crystals grown form 0.1 M succinic acid, 15% w/v PEG 3350 (left), 20% w/v PEG 3350 (middle), and 20% w/v PEG 3350, 0.2 M $MgCl_2$, 0.1 M Tris (pH 8.5) (right).

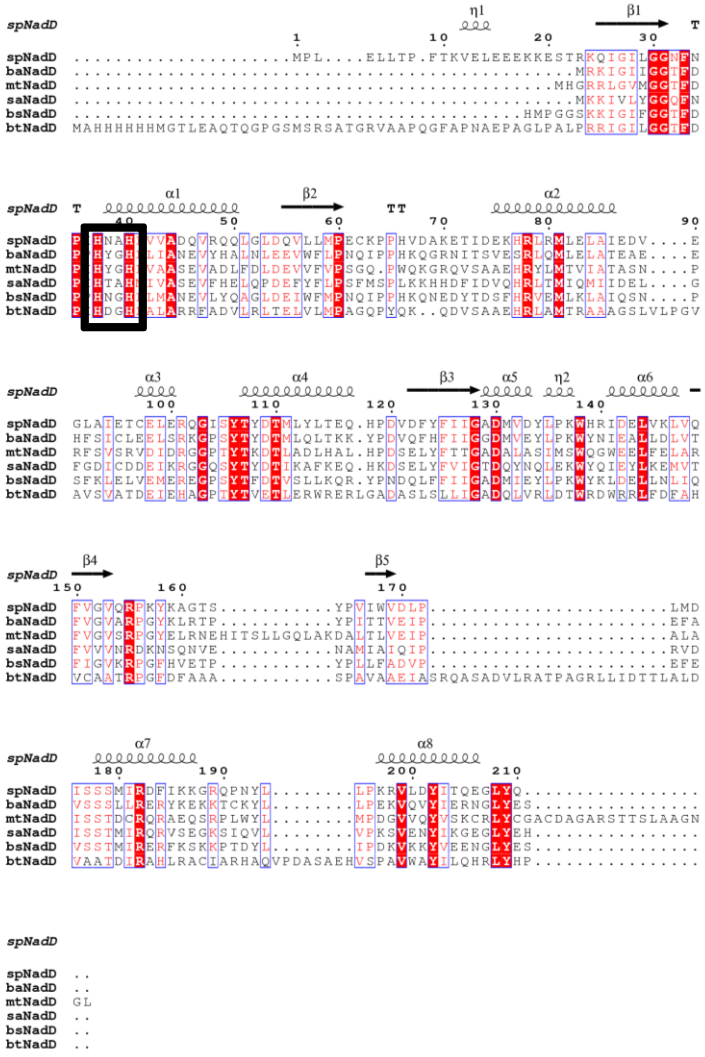


Figure 3.8: SpNadD sequence alignment. This figure shows the sequence alignment of spNadD and homologous proteins that have their structures deposited to the PDB. NadD homologs originate from the following organisms: *Bacillus anthracis* (baNadD), *Mycobacterium tuberculosis* (mtNadD), *Staphylococcus aureus* (saNadD), *Bacillus subtilis* (bsNadD) and *Burkolderia thailandensis* (btNadD). The area highlighted in black box is the location of the highly conserve (T/H) XGH motif that is characteristic of proteins in the nucleotidyl transferase superfamily (3).

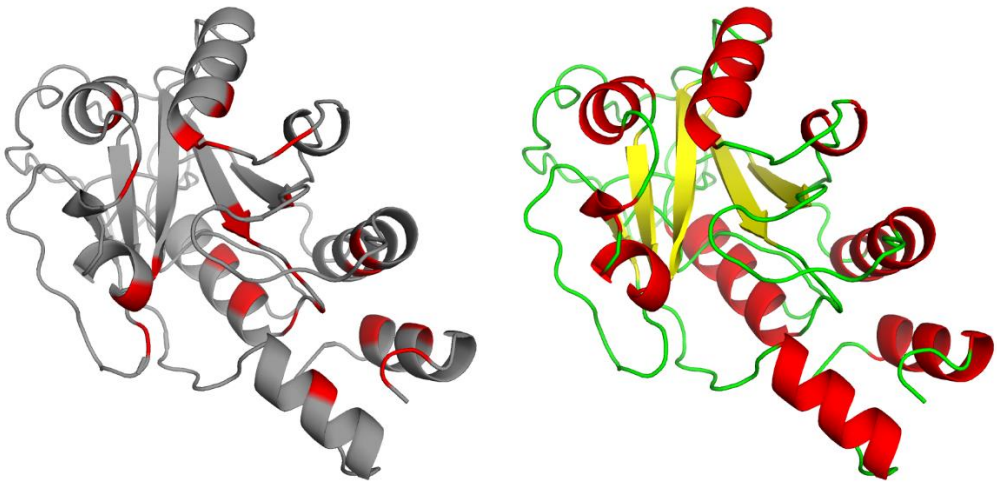


Figure 3.9: SpNadD homology model. Cartoon representations of spNadD model based on *Bacillus anthracis* NadD structure (PDB code: 3E27). Areas highlighted in red (*left*) indicate sites on the model that are considered highly conserved.

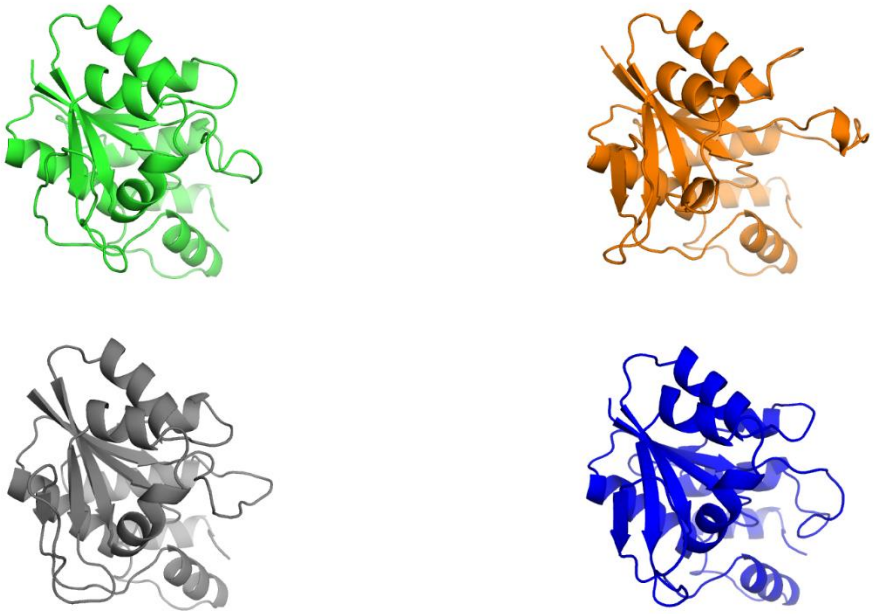


Figure 3.10: SpNadD homology models. Cartoon representations of various spNadD models. The model in green was generated using the baNadD structure (PDB code: 3E27) as a template; the orange model was generated using bsNadD structure (PDB code: 1KAM), the gray model was generated using saNadD structure (PDB code: 2H29), and the blue model was generated using paNadD structure (PDB code: 1YUN) as a template.

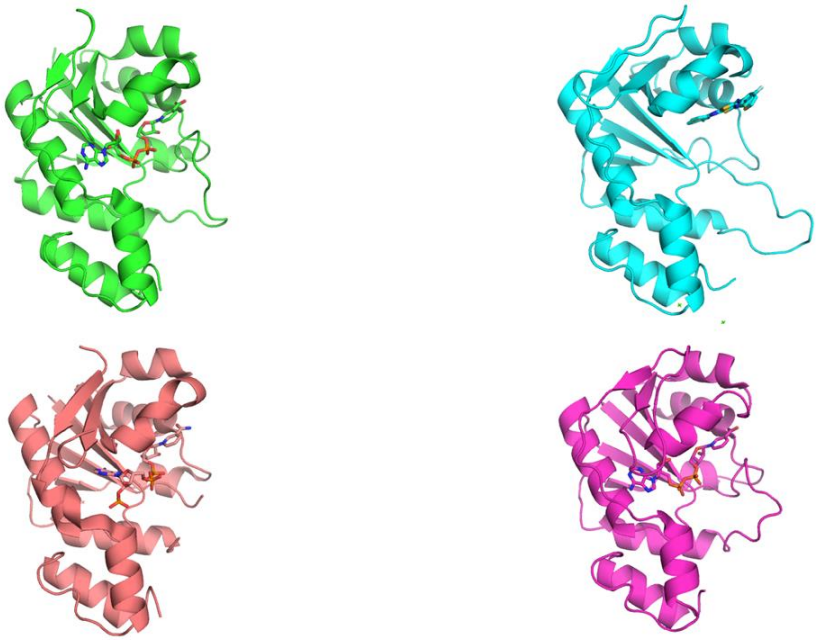


Figure 3.11: SpNadD structural homolog active site comparisons. This figures show comparisons the of active site locations on single NadD chains for proteins originating from *Staphylococcus aureus* (saNadD-green), *Bacillus anthracis* (baNadD-cyan), *Mycobacterium tuberculosis* (mtNadD) and *Bacillus subtilis* (bsNadD). Each of the structures contain a NaAD molecule within the site, except for the bsNadD structure which contains an inhibitor molecule.

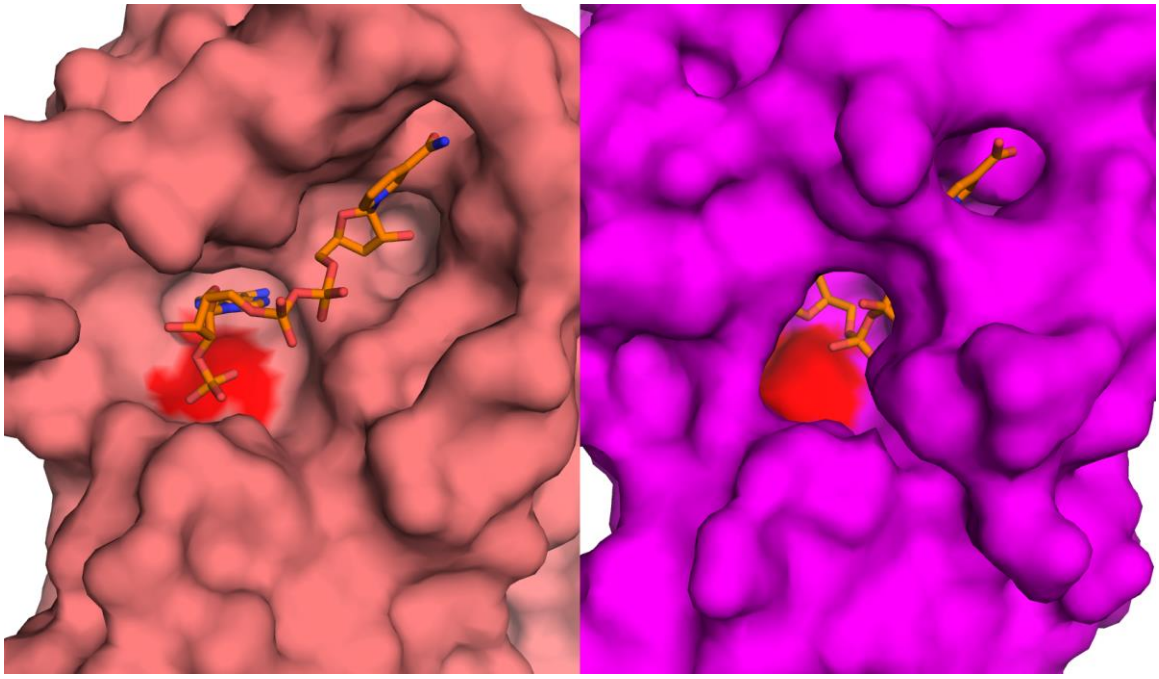


Figure 3.12: SpNadD putative substrate binding pocket. Surface representations of mtNadD (left) and bsNadD (right) showing open and closed formations of the active site in the presence of NaAD. The area highlighted in the red is the location of the (T/H)XGH motif to show its orientation relative to the adenylate moiety of NaAD.

3.5 REFERENCES

1. Sorci, L., I. K. Blaby, I. A. Rodionova, J. De Ingeniis, S. Tkachenko, V. De Crecy-Lagard, and A. L. Osterman. "Quinolinate Salvage and Insights for Targeting NAD Biosynthesis in Group A Streptococci." *Journal of Bacteriology* 195.4 (2012): 726-32. Web.
2. Mao, F., P. Dam, J. Chou, V. Olman, and Y. Xu. "DOOR: A Database for Prokaryotic Operons." *Nucleic Acids Research* 37.Database (2009): n. pag. Web
3. Aravind, L., and E. V. Koonin. "DNA Polymerase Beta-like Nucleotidyltransferase Superfamily: Identification of Three New Families, Classification and Evolutionary History." *Nucleic Acids Research* 27.7 (1999): 1609-618. Web.
4. Han, Seungil, Michael D. Forman, Pat Loulakis, Michelle H. Rosner, Zhi Xie, Hong Wang, Dennis E. Danley, Wei Yuan, John Schafer, and Zuoyu Xu. "Crystal Structure of Nicotinic Acid Mononucleotide Adenylyltransferase from Staphyloccocus Aureus: Structural Basis for NaAD Interaction in Functional Dimer." *Journal of Molecular Biology* 360.4 (2006): 814-25. Web.
5. Nureki, O., D. Vassylyev, K. Katayanagi, T. Shimizu, S. Sekine, T. Kigawa, T. Miyazawa, S. Yokoyama, and K. Morikawa. "Architectures of Class-defining and Specific Domains of Glutamyl-tRNA Synthetase." *Science* 267.5206 (1995): 1958-965. Web.
6. Olland, A.m., K.w. Underwood, R.m. Czerwinski, M.c. Lo, A. Aulabaugh, J. Bard, M.l. Stahl, W.s. Somers, F.x. Sullivan, and R. Chopra. "Structure of Bacillus Subtilis Nicotinic Acid Mononucleotide Adenylyl Transferase." (2002). Web.
7. Zhou, T., O. Kurnasov, D. R. Tomchick, D. D. Binns, N. V. Grishin, V. E. Marquez, A. L. Osterman, and H. Zhang. "Structure of Human Nicotinamide/Nicotinic Acid Mononucleotide Adenylyltransferase: Basis for the dual substrate specificity and activation of the oncolytic agent Tiazofurin." *Journal of Biological Chemistry* 277.15 (2002): 13148-3154. Web.
8. Sorci, Leonardo, Yongping Pan, Yvonne Eyobo, Irina Rodionova, Nian Huang, Oleg Kurnasov, Shijun Zhong, Alexander D. Mackerell, Hong Zhang, and Andrei L. Osterman. "Targeting NAD Biosynthesis in Bacterial Pathogens: Structure-Based Development of Inhibitors of Nicotinate Mononucleotide Adenylyltransferase NadD." *Cell: Chemistry & Biology* 16.8 (2009): 849-61. Web.
9. Zhang, Hong, Tianjun Zhou, Oleg Kurnasov, Sara Cheek, Nick V. Grishin, and Andrei Osterman. "Crystal Structures of E. Coli Nicotinate Mononucleotide Adenylyltransferase and Its Complex with Deamido-NAD." *Structure* 10.1 (2002): 69-79. Web.
10. Matthews, B. W. "Solvent Content of Protein Crystals." *Journal of Molecular Biology* 33.2 (1968): 491-97. Web.
11. Goldschmidt, Lukasz, David R. Cooper, Zygmunt S. Derewenda, and David Eisenberg. "Toward Rational Protein Crystallization: A Web Server for the Design of Crystallizable Protein Variants." *Protein Sci. Protein Science* 16.8 (2007): 1569-576. Web.
12. Altschul, S. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25.17 (1997): 3389-402. Web.
13. Altschul, S., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215.3 (1990): 403-10. Web.

14. Walter, Thomas S., Christoph Meier, Rene Assenberg, Kin-Fai Au, Jingshan Ren, Anil Verma, Joanne E. Nettleship, Raymond J. Owens, David I. Stuart, and Jonathan M. Grimes. "Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization." *Structure* 14.11 (2006): 1617-622. Web.
15. Niesen, Frank H., Helena Berglund, and Masoud Vedadi. "The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability." *Nat Protoc Nature Protocols* 2.9 (2007): 2212-221. Web.
16. Minor, Wladek, Marcin Cymborowski, Zbyszek Otwinowski, and Maksymilian Chruszcz. "HKL -3000: The Integration of Data Reduction and Structure Solution – from Diffraction Images to an Initial Model in Minutes." *Acta Crystallogr D Biol Cryst Acta Cryst D Acta Crystallogr Sect D Biol Crystallogr Acta Crystallogr Sect D Acta Crystallogr D Biol Crystallogr Acta Crystallographica Section D Biological Crystallography Acta Cryst Sect D Acta Crystallogr D* 62.8 (2006): 859-66. Web.
17. Huang, Nian, Rohit Kolhatkar, Yvonne Eyobo, Leonardo Sorci, Irina Rodionova, Andrei L. Osterman, Alexander D. Mackerell, and Hong Zhang. "Complexes of Bacterial Nicotinate Mononucleotide Adenylyltransferase with Inhibitors: Implication for Structure-Based Drug Design and Improvement." *J. Med. Chem. Journal of Medicinal Chemistry* 53.14 (2010): 5229-239. Web.

CHAPTER 4:

STRUCTURAL AND FUNCTIONAL STUDIES OF *STREPTOCOCCUS PYOGENES* NH_3-DEPENDENT NAD^+ SYNTHETASE (SPNADE)

4.1 INTRODUCTION

SpNadE is the responsible for the formation of NAD^+ by the catalyzing the last step of the QSP responsible for the conversion of NaAD and ATP into NAD^+ (Figure 4.1) (1). This enzyme is relatively well studied, and proteins from *Bacillus anthracis*, *Bacillus subtilis*, and *Escherichia coli* are characterized functionally and structurally. (2, 3, 4). Proteins originating from *Burkholderia pseudomallei*, *Deinococcus radiodurans*, *Pseudomonas aeruginosa*, and *Vibrio cholarae* have their structures determined, but there are no publications describing these enzymes.

SpNadE also belongs to the Adenosine Nucleotide α-Hydrolase (AANH) Superfamily (6, 7, 8). Proteins in this superfamily have Rossman-like alpha/beta/alpha sandwich fold (InterPro code: IPR014729) and are characterized by the presence of a "PP-loop" commonly found close to the N-terminus of the protein (8). In this loop a "SGGXD" motif is commonly found which is considered to be important for phosphate binding (8). Other proteins within this superfamily include: GMP synthases, tRNA methyl transferases, and arginosuccinate synthases (8).

Previous studies have identified NadE homologs as dimeric proteins (2, 3, 4). The dimer has two binding sites for ATP and two binding sites for NaAD (2, 4). The ATP binding site is contained within a single protein chain, while NaAD binding site corresponds to a lateral cleft stretching across both protein chains (2, 4). The formation of NAD^+ is proposed to operate in an ordered-sequential fashion with binding of ATP occurring first, followed by the binding of NaAD for the initial formation of an NaAD-adenylate intermediate (3, 4). The adenylate intermediate has been shown to be stabilized by interaction with two Mg^{2+} ions (4). Upon formation of the intermediate species, it is presumed that a nucleophilic attack by the lone pair of free NH_3 occurs at the C7 carbonyl carbon of the adenylate species (2). This attack results in the release of AMP and inorganic pyrophosphate (PP_i) leaving NAD^+ to be reduced for eventual NADH formation (2). This incidence, by the action of NH_3 only, has been identified in *E. coli* and *B. subtilis*.

NAD^+ synthetases in *Mycobacterium tuberculosis* (mtNadE), *Saccharomyces cerevisiae* (scNadE), and *Homo sapiens* (hsNadE) are glutamine dependent. Glutamine acts as a $-NH_3$ donor to promote the same nucleophilic attack as described above and leaves NAD^+, AMP, PP_i, and glutamate as its end-products. (9). HsNadE and scNadE are considered to be exclusive to glutamine while mtNadE has been shown to be able to utilize both glutamine and NH_3 (4, 9).

Rodionova *et al.* identified that the *nadE* gene is essential for bacterial survival (*10*). In knockdown studies, using *Mycobacterium smegmatis,* they showed *nadE* (or *nadD*) silencing produced diminished growth of the bacterium in culture (10). The Database for Essential Genes (DEGG) also showed that NadE, from GAS, contains

significant sequence alignment to essential, proteins produced in other pathogenic species (i.e. *S. aureus and V. cholarae*) (11).

Therefore, NadE seems to be a promising new target for development of antibacterial compounds. There is still very little information identified about this protein in GAS. This dissertation provides the first information the kinetic activity of the protein, as well as, the first structures of NadE originating from *S. pyogenes*.

4.2 Results and Discussion

Protein Production

Recombinant spNadE was successfully produced in high quantities. It was estimated that after the nickel-affinity chromatography there was, on average, 340 mg of spNadE, per 1 L culture. The last step of protein purification involved size-exclusion chromatography. As judged by the SDS-PAGE the protein purity was greater than 90%. The results from the size-exclusion step indicate the presence of one major peak corresponding to a spNadE dimer, when compared to commercial standards. These results were consistent with DSL results that indicate presence of an assembly in solution with molecular weight of 72 ±5.3 kDa. The molecular weight of a spNadE with the purification tag is 34 kDa. Attempts to remove the purification tag, with TEV protease, were unsuccessful.

NadE Kinetic Data

At increasing NaAD concentrations, spNadE shows Michealis-Menten kinetics. With spNadE at a 13 pmol concentration, results showed a K_m of 375 µM. The k_{cat} was

4.16 s^{-1} and the k_{cat}/ K_m was 11.09 s^{-1} mM^{-1} (Table 4.1). At the sample protein concentration, with increasing ATP, spNadE showed a K_m of 304 µM , a k_{cat} of 5.63, and $k_{cat}/$ Km of 16.55. At experiment concentrations of greater than 4 mM ATP, results showed ATP in excess caused an inhibitory effect. All experiments were conducted in triplicate, and these are the average values of each cuvette experiment. Kinetic results provide evidence to support that this protein is the most efficient of the QSEs. There is no other published data for spNadE kinetics and variability of results could be observed in experiments prepared *in vitro*. It also cannot be excluded that the presence of the purification tag may impact the enzymatic properties (12, 13). As mentioned previously, the his-tag removal experiments were unsuccessful.

Experimental results from protein homologs from *Bacillus anthracis* (baNadE) (58.5% sequence identity) indicate approximate Km values ($K_{m(app)}$) for NaAD and ATP are 152 µM and 289 µM, respectively; the *Bacillus subtilis* (bsNadE) homolog (56.7 % sequence identity) reported a $K_{m(app)}$, for NaAD, at 179 µM and ATP at 196 µM (2, 14). No data was given for k_{cat}/ K_m (2, 14). In comparison to our experimental results our enzyme could be regarded to have a lower substrate affinity. Refer to Table 4.1 for a report of the kinetic results for spNadE compared to homologs from *Bacillus anthracis* and *Bacillus subtilis.*

THERMAL STABILITY

DSF experiments performed in conditions with various pH and NaCl concentrations showed that the recombinant spNadE was more stable in low pH, high salt conditions (pH 4.0-4.5 and 500 mM NaCl) (Figure 4.2).

The results of the small molecules impact on protein stability indicate that the top stabilizing molecules, mellitic acid and 2-pyridine sulfonic acid, increase thermal stability of spNadE by 10 ° C as compared to the control (melting at 50 °C). Phenyl-2-thiourea and tetraamine copper (II) sulfate stabilized spNadE up to 58 °C. Co-crystallization experiments are currently being conducted to structurally confirm the binding of these molecules.

Crystal Screening and Optimization

Crystals were obtained from Crystal Screen 1 (Hampton Research, Viejo CA) optimization experiments in the following condition: 0.1 M Tris (pH 8.5), 20% w/v PEG 4000, and 0.2 M magnesium chloride. In hanging and sitting drop preparations small, needle-like clusters of crystals were obtained within 24 hours. While, these crystals were not suitable for diffraction, it was later found that underneath the needle clusters, within the same condition, were thin, rectangular, plate-like crystals (Figure 4.3). These crystals diffracted well, yet appeared inconsistently. The data obtained from the plate-shaped crystals allowed for the determination of the spNadE crystal structure to 2.5 Å and 2.1 Å.

Significant effort was placed into optimization of crystallization conditions. The optimization attempts included crystallization in various temperatures, reductive methylation of the protein (for details see Chapter 3, page 63), and use of various additives. The additives used in these experiments could be divided into three groups. The first group included spNadE substrates, products or their analogs. The second group contained compounds, i.e. glycerol, that were introduced in order to slow-down the rate of crystallization, and the third group included compounds that were increasing protein

stability. Despite the fact, that thousands of crystallization experiments were performed no significant improvement of the crystallization conditions was obtained.

ANALYSIS OF THE SPNADE SEQUENCE

Genomic analysis indicate that the reading frames for the QSEs are dispersed within the genome, and are not expressed from the same operon (5). In other pathogenic bacteria (i. e. *Staphlyococcus aureus* and *Helicobacter pylori*) the *nadE* reading frame was, also, not found to be included on an operon with genes encoding for homologous QSP proteins. (5). In GAS, when compared to the other QSEs, the *nadE* reading frame is located the farthest downstream, from the origin, at position 1288286 to position 1289110. (5).

The spNadE sequence contains 283 amino residues per protein chain (566 total in the dimer). An NCBI Protein BLAST search against sequences of proteins, with structures deposited to the PDB, revealed the presence of five major protein segments: the active site, the ATP binding site, the NAD^+ binding site, the dimer interface, and the Mg^{2+} binding site(s) (Figure 4.4) (16). In accordance with BLAST results, the ATP binding site is considered to be composed of residues: 53-61 (L**G**I**SGG**Q**DS**; amino acids marked in bold are highly conserved), 88-91, 150, 168, 197, and 217-219. The NAD^+ binding site is composed of residues: 53-56, 60-61, 89-91, 140, 144, 147-148, 150, 169, 174, 178-181, 184, 219-220, 222, 234. The residues of the dimer interface include residues: 33-34, 37, 39-41, 114-116,119-120, 123-124, 141, 145, 148-149, 151-153, 155-157, 159-160, 181-182, 185-190, 193-194, 268-270. Magnesium binding residues include residues: 56, 60, 174 and 218. Areas of highest sequence conservation are between residues 53-61 and 143-246. In NadE from GAS, there is a "SGGQD" motif, which corresponds to residues 56-

60. It should be noted that BLAST results do show residue overlap between segments. Substrate-bound spNadE structures will need to be determined to confirm this overlap.

Sequence alignment studies comparing spNadE and hsNadE, using Clustal Omega (15, 16), reveal that the identity between these sequences is approximately 24% (Figure 4.5). The low similarity between these proteins gives confidence that spNadE would be an optimal template for structure-based drug design.

STRUCTURAL ANALYSIS OF SPNADE

OVERALL STRUCTURE

The NadE structure bound to Mg^{2+} and SO_4^{2-} ions ($spNadE_{sulf}$) was the first of the NadE structures to be determined (Figure 4.6). The protein crystallized in primitive orthorhombic system and $P2_12_12_1$ space group with two protein chains in the asymmetric unit. The structure was determined at 2.5 Å resolution. An NadE homolog from *Salmonella typhimurium* (PDB code: 3HMQ) was successfully used as the starting model for molecular replacement. The structure of $spNadE_{sulf}$ was refined to 0.171 R_{cryst} and 0.225 $R_{free.}$ The model of the structure includes a dimer and the protein chains include residues 9-214 and 229- 282. Residues 215-228 from the highly conserved active site loop, are missing in both structures, which indicates that they are disordered in the absence of substrates. Similarly, the polyhistidine tags were also not modeled, as there is no visible electron density that corresponds to this fragment of the molecule. It is assumed that high entropy of the segment was the cause of in ability to visualize the electron density through x-ray crystallography.The model of $NadE_{sulf}$ was deposited to the PDB with accession code 5HUH.

The $NadE_{sulf}$ structure was used as a starting model during determination of the NadE apo structure (NadE-apo). The NadE-apo structure was determined at 2.1 Å resolution, and the final model was refined to R_{cryst} value of 0.168 and an R_{free} of 0.212 (Figure 4.7). The model included residues 9-96, 98-215, 228- 282 for each of the protein chain forming the dimer present on the asymmetric unit. The model of NadE-apo was deposited to the PDB with accession code 5HUJ.

The single chain of spNadE has primarily alpha helical character with 11 alpha helices (68 % of secondary structure), 3 parallel beta strands (18%), and 2, 3_{10} helices (13 %) (Figure 4.8, Figure 4.9). The sequence and structure based searchers revealed that there are several spNadE homologs that have their structure experimentally determined. Homologous NAD^+ synthetases that had the highest sequence identity originated from *Salmonella typhimurium* (PDB code: 3HMQ, 68 % sequence identity; stNadE), *Escherichia coli* (PDB code: 1WXI, 65% sequence identity; ecNadE), *Bacillus anthracis* (PDB code: 2PZA; 58% sequence identity; baNadE) and *Bacillus subtilis* (PDB code: 2NSY; 57% sequence identity; bsNadE) (18). Structure of these proteins were used to identify putative substrate binding sites for spNadE, and gave the following RMSD values, using Coot SSM superpose: 0.69 Å (over 251 Cαatoms) for stNadE; 0.62 Å (over 257 Cα atoms) for ecNadE, 0.74 Å (over 254 Cα atoms) for baNadE, and 0.85 Å (over 254 Cα atoms) for bsNadE. Figure 4.10 highlights the regions of the spNadE structure that are the most highly conserved, based on sequence alignments from these homologs.

QUATERNARY STRUCTURE

Gel-filtration and DSL results indicated that spNadE forms dimers in solution. These results are in agreement with structural analyses that show the presence of a dimeric assembly in the asymmetric unit. We assume that the dimer observed in the crystal corresponds to the biological assembly of spNadE. According to PDBpisa calculations (19.), the dimer interface for spNadE-apo covers 2407 $Å^2$ and the same interface for $spNadE_{sulf}$ covers 2407 $Å^2$.

ACTIVE SITE

Sequence homology analysis, using NCBI Protein BLAST, in combination with structural studies, indicate that there are approximately 18 residues responsible for the architecture of the active site. Through analysis of the structure it was evident that there was no electron density for 15 of the residues due to possibly high entropy in that region. Through sequence homology studies of this region, specifically residues 215-223, it was determined that these residues were highly conserved and make up a loop structure within that region. Using a structural homolog from bsNADE (PDB code: 1EE1) we modeled the loop into this region in order to obtain some impression of the activity of this loop within this region (Figure 4.11). It should be noted that bsNadE (PDB code: 1EE1) structure was one of the few structures that determined the orientation for the loop structure (20). This loop has also been indicated in the stabilization of Mg^{2+} within the ATP binding site, as the ATP binding site shares 3 residues with this site. The architecture of this site will be discussed in greater detail in the next section. The residues for the loop of interest include E215, K216, V217, P218, T219, A220, D221, L222, E223 from spNadE. In bsNadE

residues include: L222, K223, E224, P225, T226, A227, D228, L229, and L230. Both loops are have approximately 67% sequence identity.

SpNadE ATP/ Mg^{2+} Binding

ATP and Mg^{2+} binding sites in NadE were shown to be composed of residues from a single protein chain. Sequence homology studies of structural homologs of the spNadE provide evidence to support that binding of the ATP occurs solely within the individual chains of the dimeric assembly and not at the interface. Residues involved in the conserved "SGGXD" motif, described above, are also is found at this site.

Sequence homology studies of spNadE identifies residues: L53, G54, I55, S56, D60, S61, V89, R90, L91, R150, T168, K197, V217, P218, and T219 to be responsible for the providing a binding pocket for ATP. Residues P218 and T219 (as well as any other residues between 216-227) are not visible in the electron density due to a conformational flexibility in that region. To describe the architecture of this site, I used structures from ecNadE (PDB code: 1WXI) (Figure 4.12, 4.13) and bsNadE (PDB code: 1EE1) (Figure 4.14).

Sequence homology results also show that the residues responsible for Mg^{2+} binding are the most highly conserved (100% sequence identity) between the structural homologs. BLAST results for spNadE show Mg^{2+} binding residues include: S56, D60, E173, and T218 (T218 was not visible, as described above) (Figure 4.15, 4.16). (18).

The ATP binding sites between the spNadE-apo and ecNadE AMP-bound structure (PDB code: 1WXI) do not show any significant cavity differences. When the bsNadE (PDB code: 1EE1) was superposed on the structure it provided evidence that the conserved loop

region closes off the ATP binding site and could play a role in the prevention of new ATP molecules entering into the site (Figure 4.17). We assume that when NAD^+ is released, the site re-opens and allows for access of another ATP molecule.

NAAD BINDING SITE

According to previous studies on similar -NH_3 dependent NadE homologs, NaAD has been shown to bind across the dimeric chains in a cleft that is below the ATP binding site (Figure 4.18) (20). This binding event should allow for the interaction of the γ-phosphate of ATP with the carboxyl group of the pyridine ring of NaAD in order to allow for the formation of the NaAD-adenylate intermediate. According to sequence homology, residues responsible for the formation and stabilization of NaAD binding site in spNadE include: L53- S56, D60, S61, V89-L91, F140, G143, N144, A147, R148, R150, T168, E173, F178-K181, D184, T219, A220, L222, and A234 (Figure 4.19) (18). SpNadE residues T219, A220, and L222 are not visible in the electron density. When compared to bsNadE, NaAD binding site residues have approximately 85% residue identity.

In the surface representation comparisons between apo spNadE and, NaAD-bound, bsNadE it was identified that the substrate binding pocket, for bsNadE, was visually, larger than the predicted NaAD binding site for spNadE. Using Coot and PyMOL, we superposed the structures and measured the distance between two residues that were directly across from each other, on opposite sides of the binding pocket. These measurements were taken across the centermost point of the pocket in an attempt to confirm an actual size difference. More specifically, we, first, measured the distance between the ζ carbon of F128 and the δ sulfur of M253 of bsNadE. Then, we measured the distance between the ζ carbon of F139

and δ carbon K264 of spNadE. Results from these residue measurements identified that the distance between F128 to M253 in (bsNadE) was 1.8 Å wider than the distance between F139 and K264 (spNadE) (Figure 4.20). It was concluded that NaAD binding site is larger when the substrate is bound.

Moreover, further analysis of the area shows that NaAD binding site expansion seems to occur as an effect of the closure of the ATP binding site, after ATP is bound. This occurrence is presumed to be similar to how the compression one end of a balloon causes expansion of the opposite end. In contrast, ATP site expansion was identified when NaAD is unbound. When NaAD is not bound, the structure reveals an open conformation of the ATP binding site is and a more narrow conformation NaAD site, as described above (Figure 4.21). This observation lead us to believe that the general mechanism for NAD^+ biosynthesis is an ordered-sequential reaction (21) (Figure 4.22). We believe that the ATP binding event occurs first, then the loop structure of the active site clamps down, which promotes the open conformation of the NaAD site for acceptance of the substrate (Figure 4.21). More kinetic and structural studies will need to be done to confirm this occurrence.

NH_3 CHANNEL

NAD^+ biosynthesis can also occur through the utilization of glutamine as an $-NH_3$ donor, for nucleophilic attack, to convert the NaAD-adenylate intermediate into NAD^+. This mechanism is most common in *Mycobacterium tuberculosis* (*M. tuberculosis*), *Homo sapiens*, and *Saccharomyces cerevisiae* (4). MtNadE has been identified to have the ability to utilize $-NH_3$ and glutamine to make NAD^+ (4, 22).

The mtNadE (PDB code: 3SYT) structure is a large (600 kDa) octomeric biological assembly structure with general characteristics similar to a cross, with a hollow, center cavity (9). At the apex of each corner of the cross resides the C-terminal NAD^+ synthetase domain (9). This domain is a structural homolog of spNadE (RMSD of 1.6 Å) (Figure 4.23). Beneath the synthease resides glutaminase, an N-terminal amidotransferase, region that that promotes the loss of $–NH_3$ necessary for the NAD^+ synthetase reaction (Figure 4.24) (9). From work conducted by LaRonde-LeBlanc *et al*., it was identified that glutamine enters through a "tunnel system" in which ends at a Glu-Lys-Cys triad that is proposed to promote the loss of $-NH_3$ from the δ carbon of glutamine (Figure 4.25) (9). Once $-NH_3$ is released it passes into a second "tunnel system" to initiate the nucleophilic attack on the NaAD adenylate intermediate, as described previously. The entire occurrence is proposed to travel approximately 73 Å before the reaction is complete (Figure 4.26, 4.27). K_{cat} for this reaction was 0.68 and K_{cat}/K_m was defined at 0.4 s^{-1} mM^{-1} (9).

Structural homology studies have lead us to conclude that the direction of nucleophilic attack comes from underneath the spNadE structure through channels formed at the interface of the A and B chains (Figure 4.28). The residues that make up this interface are made up of loop structures and include: Q18, L19, G167, T168, D169, N174, F179, T180, K181, F182, G183, D184, G185, G186, A187, D188, F193, R270, H271, L272, P273, and I274. This area is not highly conserved, and contains few hydrophobic residues. This observation could provide reasoning for ease of access to shared NAD^+ binding site residues within this region. We infer, from this result, that passage of $-NH_3$ in spNadE will occur at the same entrance point that the tunnel system affords for mtNadE synthetase region. Further tests will need to be conducted to confirm this assumption.

4.3 Materials and Methods

Protein Production and Crystallization:

The methods used for protein production and crystallization were the same as those described in Chapter 2, pages 27-30.

Dynamic Light Scattering:

The methods used for Dynamic Light Scattering are outlined in Chapter 2, page 30.

Thermal Stability Tests:

The methods used for Thermal Stability test are outlined in Chapter 2, page 30-31. Small molecule stability screening tests were done utilizing the strategy outlined in Chapter 3, page 64.

NadE Kinetic Assay

NadE activity was conducted at 1 mL, using a quartz cuvette. Activity reaction solution concentrations are as follows: 60 mM HEPES (pH 7.5), 2 mM NaAD, 2 mM ATP, 10 mM ammonium chloride, 10 mM $MgCl_2$, and 20 mM KCl, 0.5 µg of NadE, 1% ethanol and 1 µg of alcohol dehydrogenase (ADH; Sigma-Aldrich St. Louis, MO) (protocol was adapted from Nessi *et al.*) (11). ADH was necessary to confirm NAD^+ formation through its ability to promote a hydride shift to form NADH. NADH formation was detected by UV at 340 nm. Reactions were measured at 340 nm for 5 minutes. Kinetic experiments were prepared with increasing amounts of one substrate while keeping the other substrate at activity reaction concentrations. ATP kinetic experiments concentrations are as follows: 0.5 mM, 1 mM, 2 mM, 4 mM, 6 mM, 8 mM, 10 mM, and 12 mM. NaAD kinetic experiment

concentrations are as follows: 0.125 mM, 0.250 mM, 0.500 mM, 1 mM, 1.5 µM, 2 mM, and 2.5 mM. Reactions were conducted using the Nanodrop 2000c UV spectrometer.

DATA COLLECTION AND DATA PROCESSING

Prior to data collection $NadE_{sulf}$ crystal was transferred to the solution of cryo-protectant composed of 10 % glycerol, 5% w/v PEG 3350, and 0.1 M ammonium sulfate. The NadE-apo crystal was cryo-cooled by direct transfer from the drop to liquid nitrogen. Diffraction data were collected at the Advanced Photon Source at Argonne National Laboratory in Lemont, IL. The data from NadE-apo crystal was collected at 19 ID of the Structural Biology Center Collaborative Access Team (SBC-CAT) and data from the $NadE_{sulf}$ crystal was collected at 21 ID-G at the Life Sciences Collaborative Access Team (LS-CAT). Diffraction experiments were performed at 100 K. Data processing was performed with HKL-2000 and the details on the diffraction experiments and collection statistics are summarized in Table 4.2.

STRUCTURE DETERMINATION, REFINEMENT AND VALIDATION

HKL-3000 in combination with MOLREP were used for structure determination. Both of the NadE structures were solved by molecular replacement. The structure of $NadE_{sulf}$ was solved first using as the starting model the NAD^+ synthetase *Salmonella typhimurium* (PDB code: 3HMQ; 67% of sequence identity). The structure of $NadE_{sulf}$ was used as a starting model for determination of NadE-apo structure. Both structures were refined with HKL-3000 and Refmac. Coot was used for model building and adjustment, as well as for structure validation. The final validation was performed with Molprobity and CheckMyMetal. Both structures together with structure factors were deposited to the

Protein Data Bank under the accession codes 5HUJ and 5HUH for NadE-apo and $NadE_{sulf}$, respectively. Information on model refinement and validation was summarized in Table 4.3.

OTHER COMPUTATIONAL METHODS

The other computational methods used are the same as were described in Chapter 2, page 33.

4.4 TABLES AND FIGURES

Table 4.1: Kinetic results from spNadE experiments. These results are compared to results reported in the literature for baNadE and bsNadE (2, 14).

Enzyme	Variable Substrate	Fixed Substrate	$K_{m(app)}$ (μM)	K_{cat} (s^{-1})	K_{cat}/K_m (s^{-1} mM^{-1})
spNadE	NaAD	ATP	375 ± 0.12	4.16	11.09
	ATP	NaAD	340 ± 0.03	5.63	16.55
bsNadE	NaAD	ATP	152	NR	NR
	ATP	NaAD	289	NR	NR
baNadE	NaAD	ATP	179	NR	NR
	ATP	NaAD	196	NR	NR

Table 4.2: Summary of NadE and $NadE_{sulf}$ data collection and processing.

Protein	$NadE_{sulf}$	NadE-apo
PDB accession code	5HUH	5HUJ
Diffraction source	APS (21 ID)	APS (19 ID)
Wavelength (Å)	1.000	0.979
Detector	MX300HS	ADSC Q315r
Crystal-to-detector distance (mm)	300	300
Rotation range per image (°)	1	1
Total rotation range (°)	200	200
Exposure time per image (s)	1	1
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
a, b, c (Å)	49.521 92.104 128.569	49.786 92.964 128.355
Resolution range (Å)	50.00- 2.50 (2.54-2.50)	39.37- 2.10 (2.14-2.10)
σ Cutoff	-3 σ	-3 σ
Total No. of reflections	19874 (856)	32708 (1586)
Completeness (%)	92.6 (82.9)	91.8 (90.5)
Redundancy	6.6 (5.3)	5.6 (5.0)
$\langle I/\sigma(I)\rangle$	7.5 (2.4)	11.9 (2.3)
$R_{r.i.m.}$	0.159 (0.588)	0.156 (0.862)
$R_{p.i.m.}$	0.060 (0.238)	0.064 (0.373)
Overall *B* factor from Wilson plot ($Å^2$)	26.8	29.6
CC ½	(0.873)	(0.645)

Table 4.3. Summary of spNadE and spNadE$_{sulf}$ structure refinement statistics

Protein	NadE$_{sulf}$	NadE-apo
Resolution range (Å)	43.72-2.50 (2.56-2.50)	39.37- 2.10 (2.15-2.10)
Completeness (%)	93.10 (92.1)	91.66 (89.5)
σ Cut-off	0	0
No. of reflections, working set	18796	31023
No. of reflections, test set	1338	1630
Final R_{cryst}	0.171 (0.222)	0.168 (0.205)
Final R_{free}	0.225 (0.279)	0.212 (0.282)
No. of non-H atoms	4301	4291
Protein	4025	4025
Ligand	27	1
Solvent	249	265
R.m.s deviations		
Bonds (Å)	0.012	0.016
Angles (°)	1.4	1.6
Average B factors (Å^2)		
Protein	32	29
Ligand	74	38
Ramachandran plot		
Most favored (%)	98.8	99.2
Allowed (%)	99.8	99.6
Visible Residues		
Chain A	9-215, 228- 282	9-97, 99-215, 228- 282
Chain B	9-215, 228- 282	9-97, 99-215, 228- 282

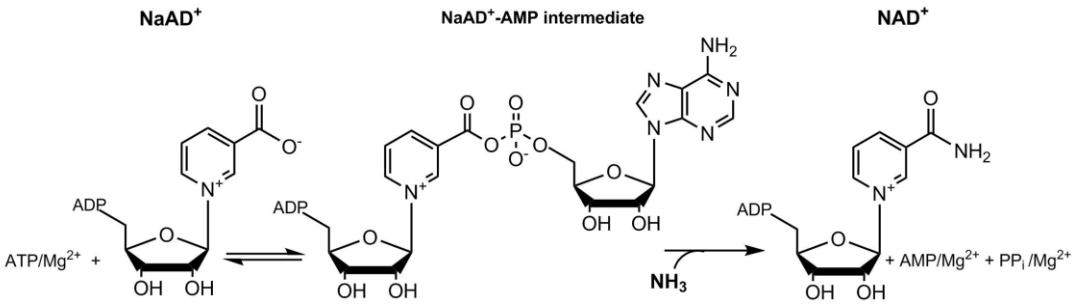


Figure 4.1: Schematic of general spNadE reaction. This diagram shows the general mechanism of an NH_3-dependent synthetase.

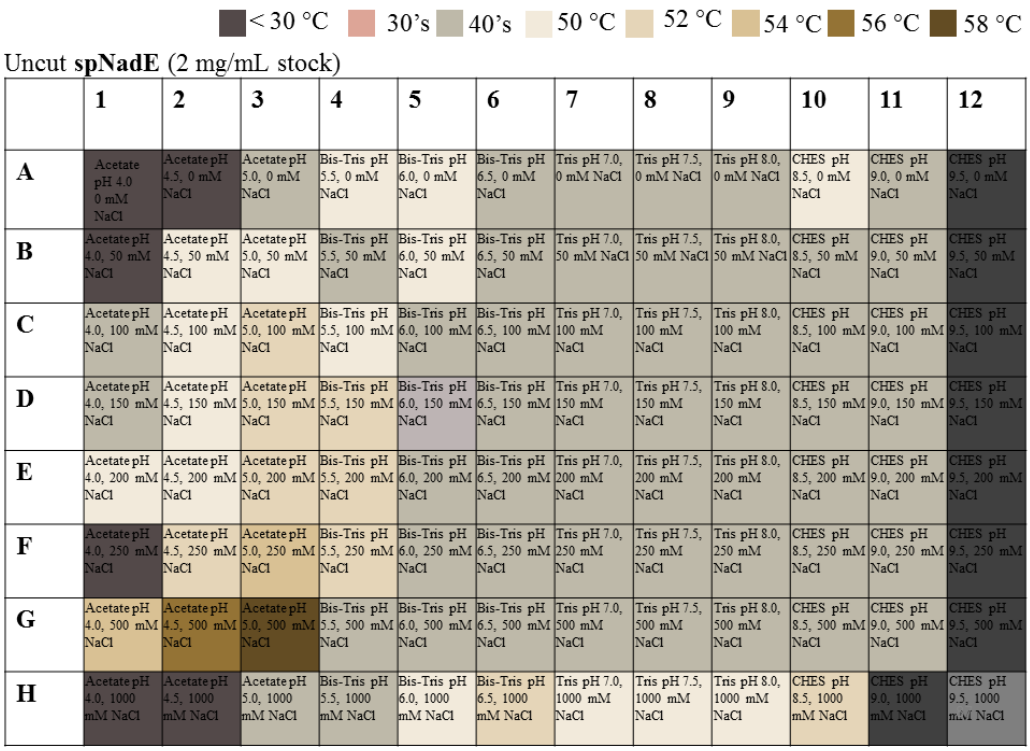


Figure 4.2: SpNadE DSF result. Chart depicts differential scanning fluorimetry results for NadE; results show that NadE has a greater thermal stability in acidic conditions.

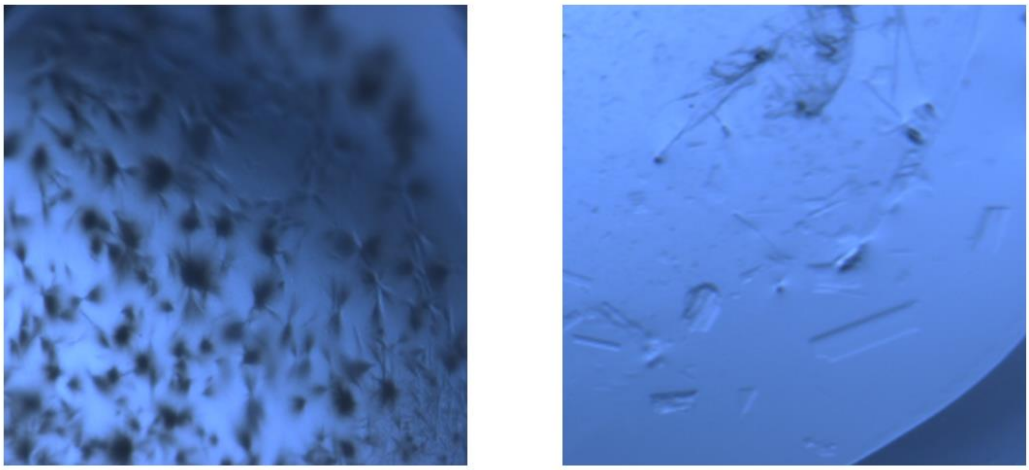


Figure 4.3: SpNadE crystallization results. The image on the *left* shows a *p*icture of the needle-like cluster of spNadE crystals. The right image shows the rectangular, plate-like crystals found underneath the needle clusters that were optimal for diffraction.

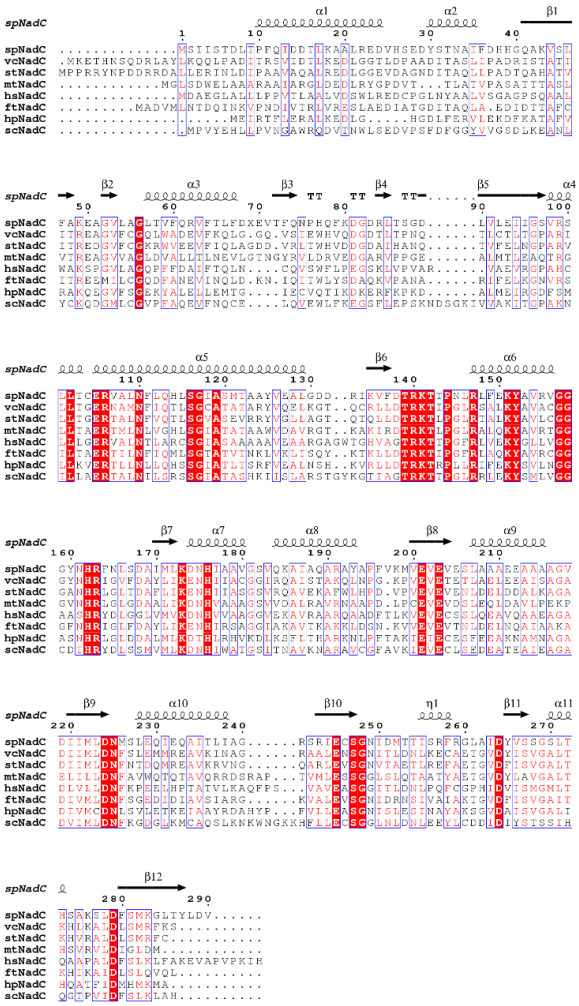


Figure 4.4: SpNadE sequence alignment. This image show sequence alignment results for spNadE and its homologs originating from *Salmonella typhimurium* (stNadE), *Eschericia coli* (ecNadE), *Bacillus anthracis* (baNadE), *Vibrio choloera* (vcNadE), *Pseudomonas aeruginosa* (paNadE), *Bortadella pertussis* (btNadE), *Fracisella tulerensis* (ftNadD), *Camphylobacter jejuni* (cjNadE), and *Helicobacte r pylori* (hpNadE)

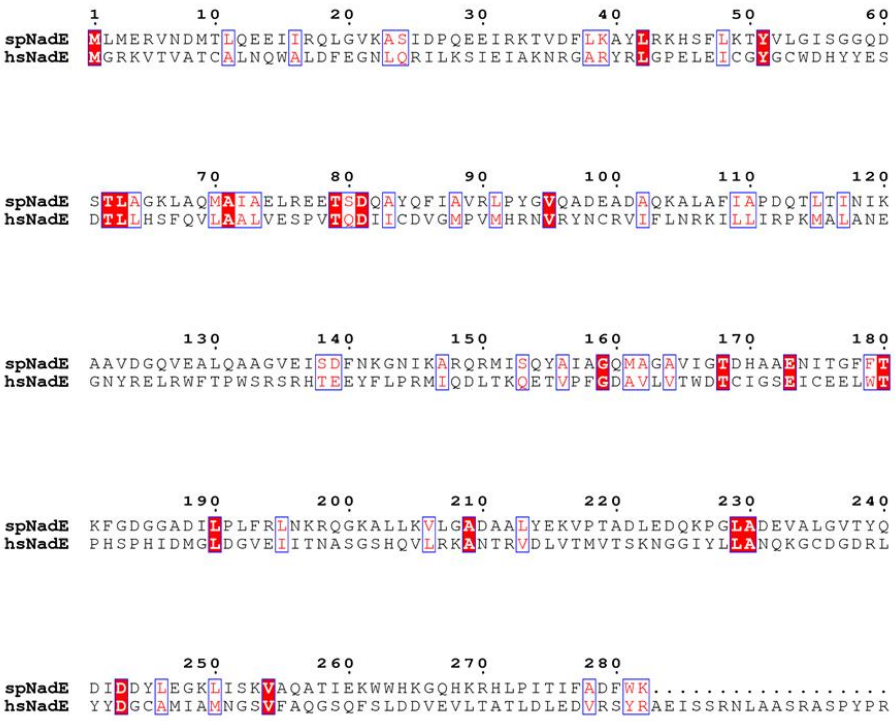


Figure 4.5: Sequence alignment of spNadE and hsNadE. This alignment result shows low sequence identity and similiarity between the bacterial and human protein. The sequence of the C-terminal glutaminase domain of hsNadE was omited.

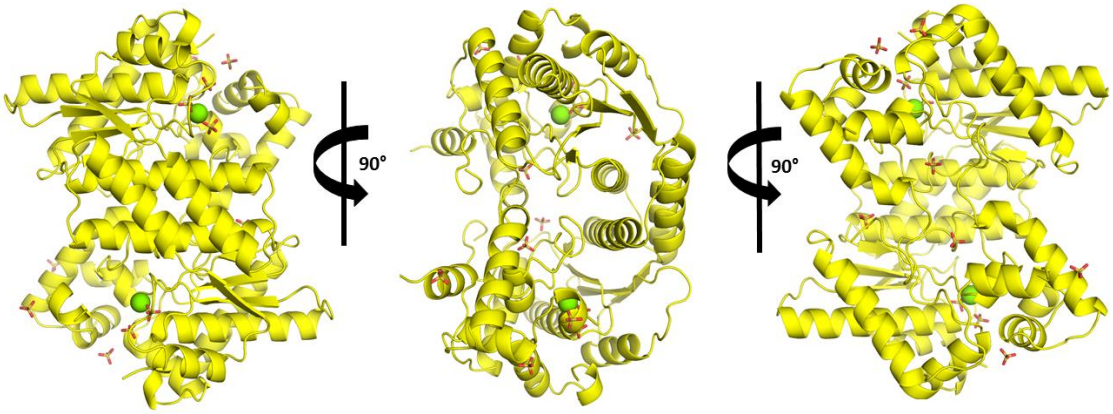


Figure 4.6: Cartoon representation of spNadE$_{sulf}$. This is a cartoon representation of spNadE$_{sulf}$ in various orientations bound to sulfate and Mg^{2+} (green spheres).

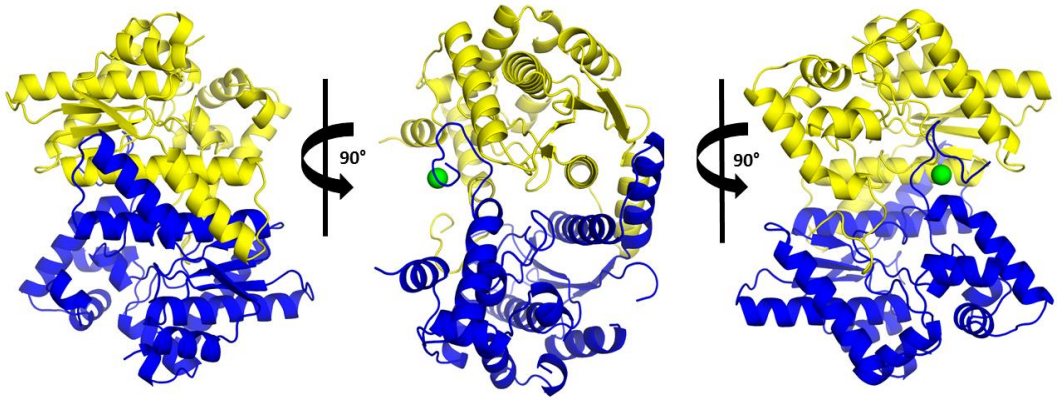


Figure 4.7: Cartoon representation of apo spNadE. This is a cartoon representation of the apo form of spNadE in various orientations. The sphere highlighted in green represents Cl^-.

Figure 4.8: The spNadE monomer. Cartoon representation of spNadE monomer at various orientations. Alpha helices are colored in cyan, beta-strands are colored red, and loop regions are colored magenta.

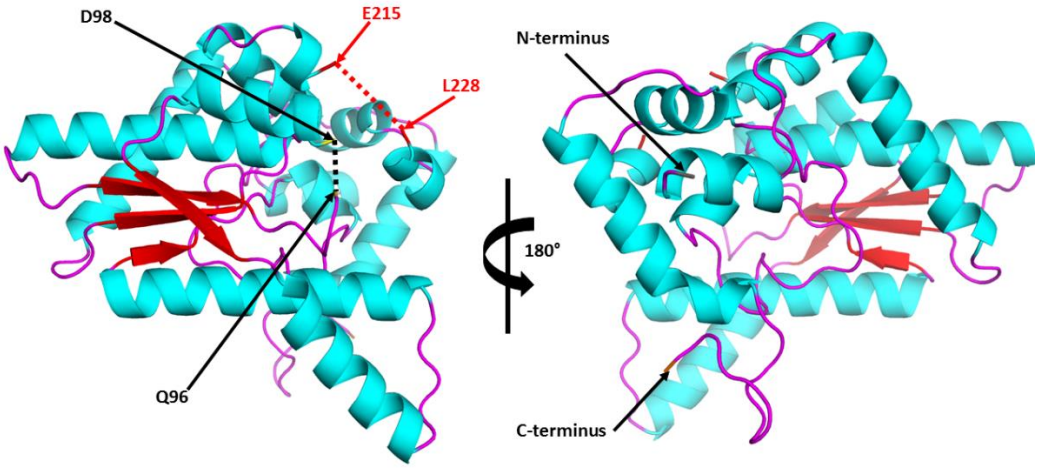


Figure 4.9: SpNadE monomer with marked missing residues. Cartoon representation of spNadE monomer shown at two different orientations. Missing residues on this structure include residues 97 and 216-227. The missing region from E215-L228 is highlighted in red to indicate that this region is highly conserved and will be described in greater detail in later figures. Alpha helices are colored in cyan, beta-strands are colored red, and loop regions are colored magenta. The N and C termini are both labeled on the structure on the right.

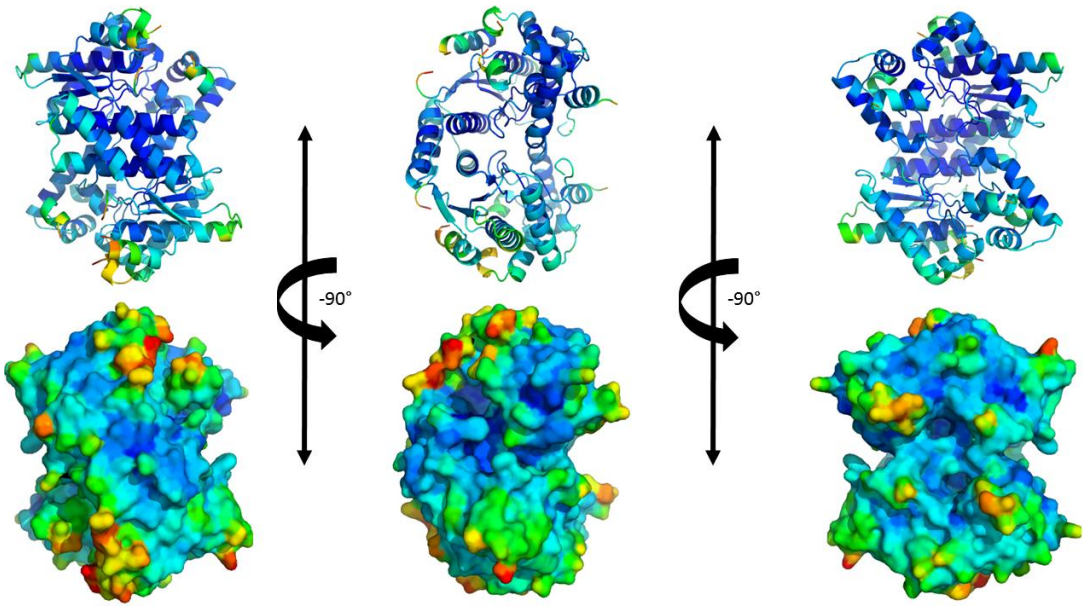


Figure 4.10: Cartoon representation of conserved regions of spNadE. Cartoon and surface representations of the spNadE dimer in various orientations. The dark blue color shows regions of structure that are highly conserved. Areas in red are the least conserved.

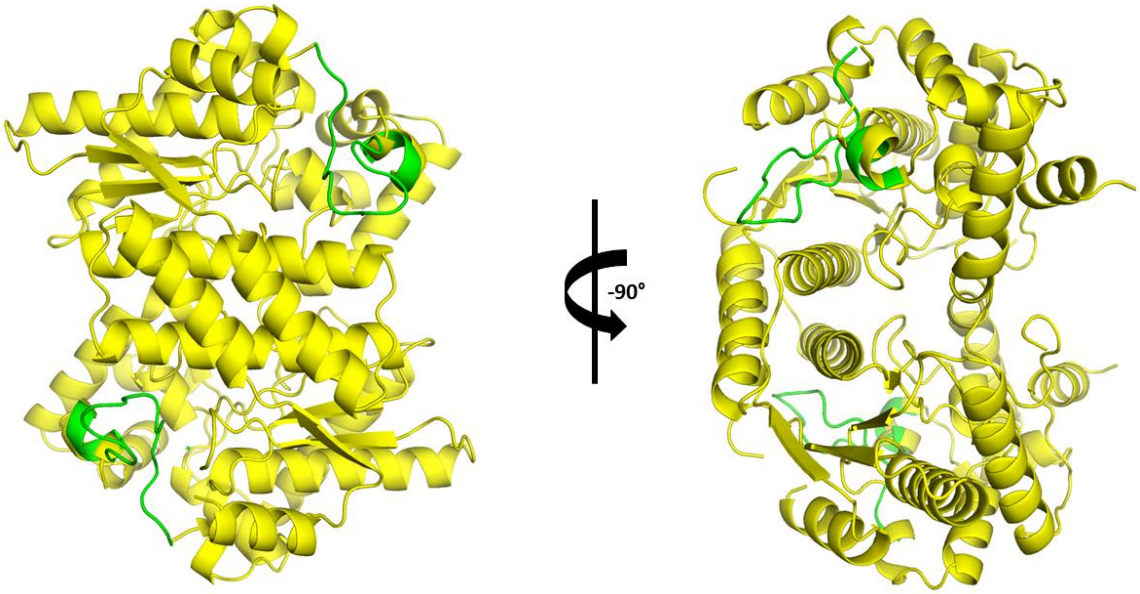


Figure 4.11: SpNadE and proposed active site. Cartoon representation of spNadE with superposed, conserved loop region (K225-L231) from bsNadE (PDB code: 1EE1).

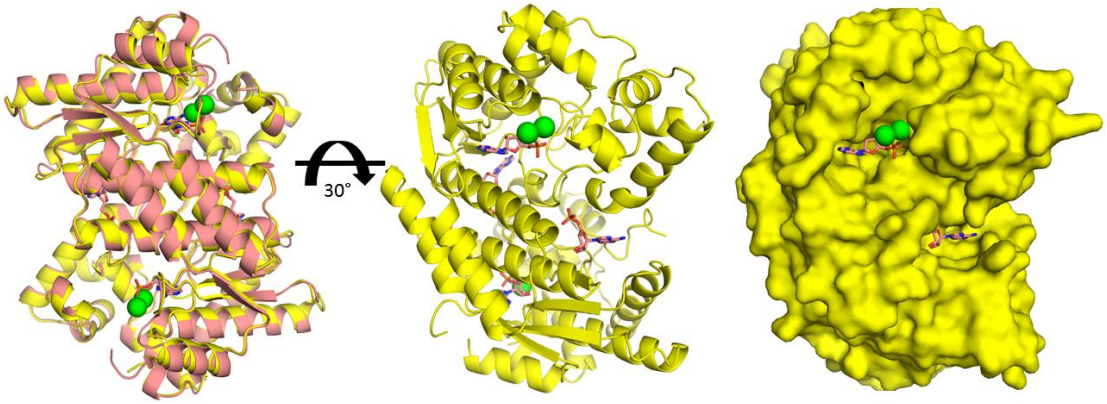


Figure 4.12: Proposed spNadE ATP binding site. Cartoon representation of spNadE superposed with ecNadE (salmon) (PDB code: 1WXI) to show probable location of ATP binding site without the conserved loop. AMP is bound to ecNadE.

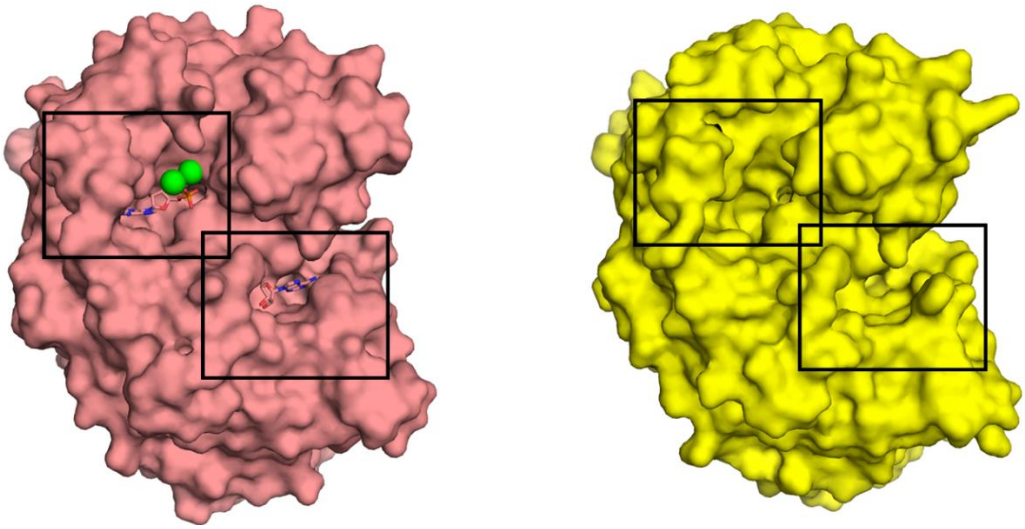


Figure 4.13: Proposed ATP binding pocket. Cartoon representation of spNadE and ecNadE (salmon) (PDB code: 1WXI) to show probable location of ATP binding site without the conserved loop. AMP is bound to ecNadE and the green spheres are two Mg^{2+} molecules. The AMP molecule bound in lower right quadrant of the ecNadE molecule may give insight into the directionality of NaAD binding within the site.

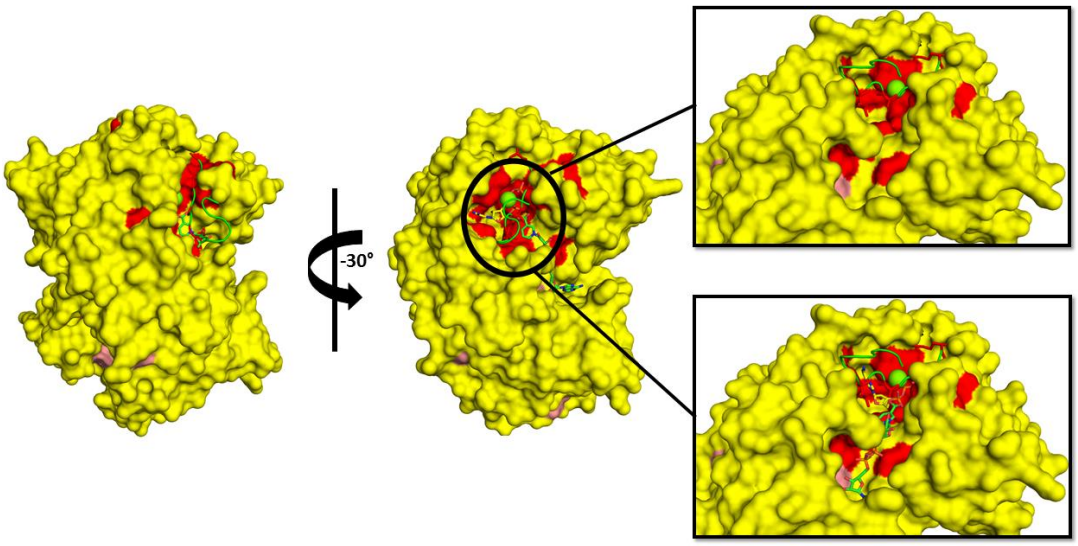


Figure 4.14: SpNadE modeled with active site loop. Surface representation of spNadE with a superposed loop from bsNadE. Areas highlighted in red represent conserved regions for the ATP and NaAD binding sites. Images to the far right are meant to display conserved residues of the loop that interact with the site without (top) and with (bottom) substrate. ATP identified in this figure is also superposed from the bsNadE (PDB code: 1EE1) structure.

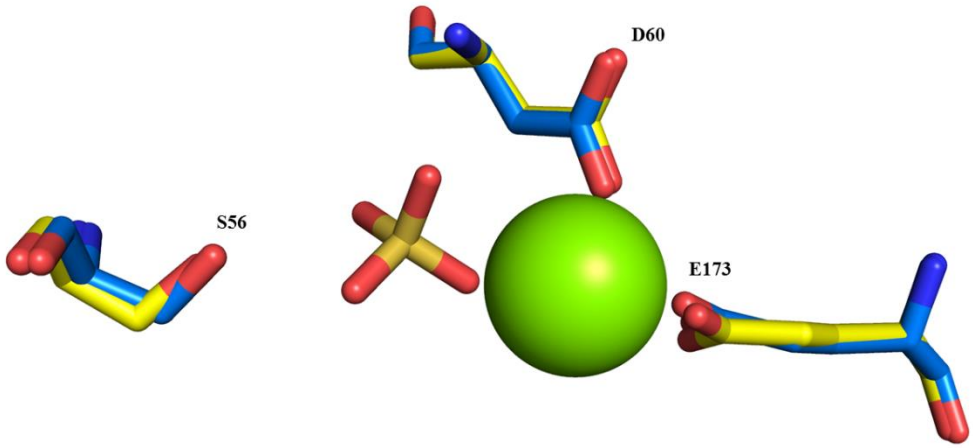


Figure 4.15: Highly conserved spNadE Mg^{2+} binding site. The superposition of Mg^{2+} binding site residues for spNadE apo structure (blue) and $spNadE_{sulf}$ (yellow).

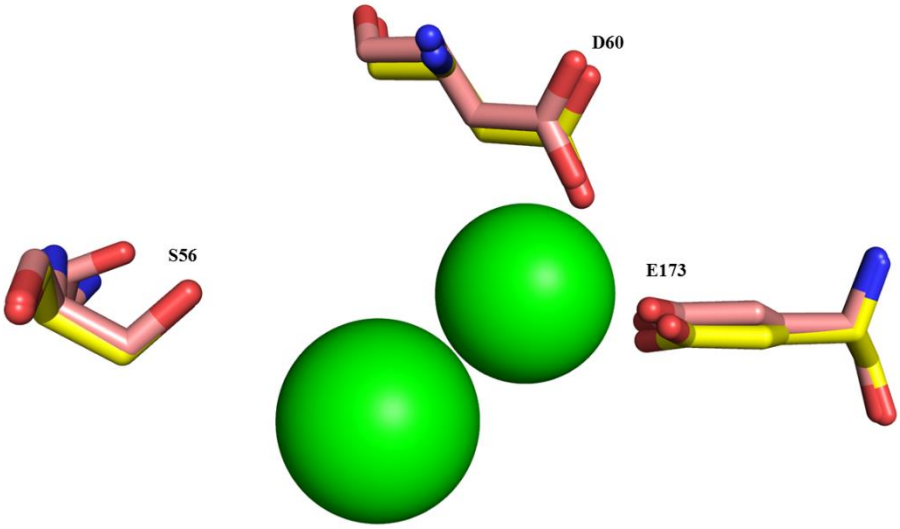


Figure 4.16: Conserved Mg^{2+} binding site superposed with equivalent fragment of ecNadE. The superposition of Mg^{2+} binding site residues for spNadE apo structure (yellow) and ecNadE (PDB code: 1WXI; salmon).

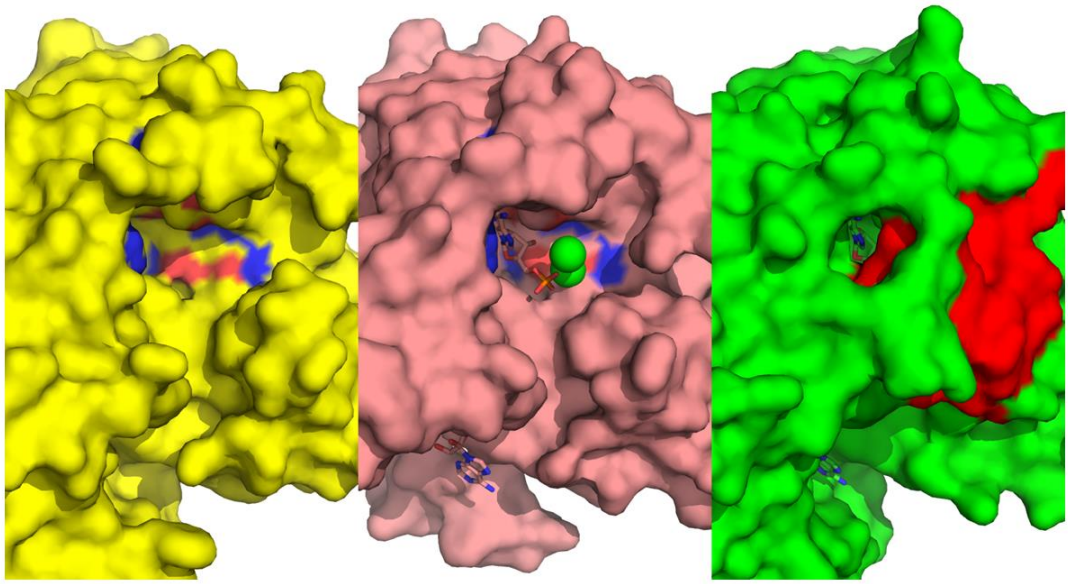


Figure 4.17: Surface representation of ATP binding site with and without the active site loop. The image to the left (yellow) is spNadE, the middle image (salmon) is ecNadE (PDB code: 1WXI), the right figure is the bsNadE (PDB code: 1EE1). For the bsNadE figure the area highlighted in red represents the conserved loop region that, essentially, closes off the binding site upon binding of ATP.

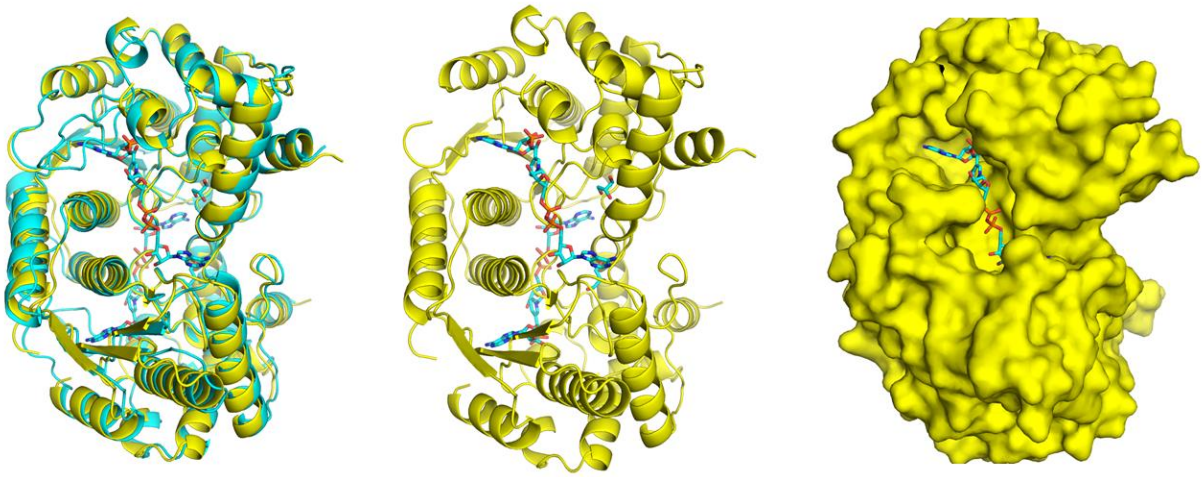


Figure 4.18: Proposed spNadE NaAD binding site. Cartoon and surface representation of proposed the spNadE NaAD binding site of NaAD from superposed bsNadE structure (PDB code: 2NSY). Notice the binding of NaAD across the interface of the dimer.

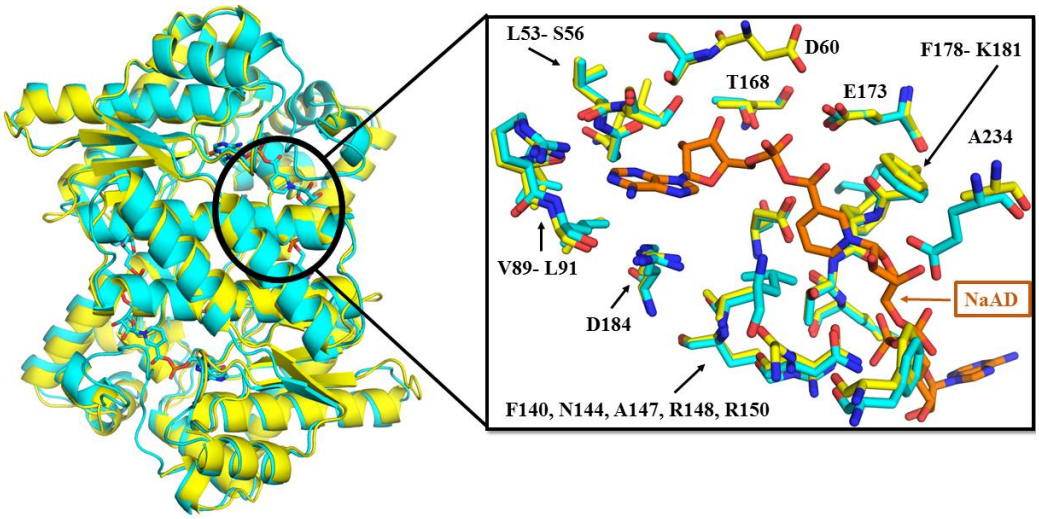


Figure 4.19: Proposed spNadE residues for NadE binding. Cartoon representation of superposed spNadE (yellow) and bsNadE (cyan) (PDB code: 2NSY) NaAD binding site residues.

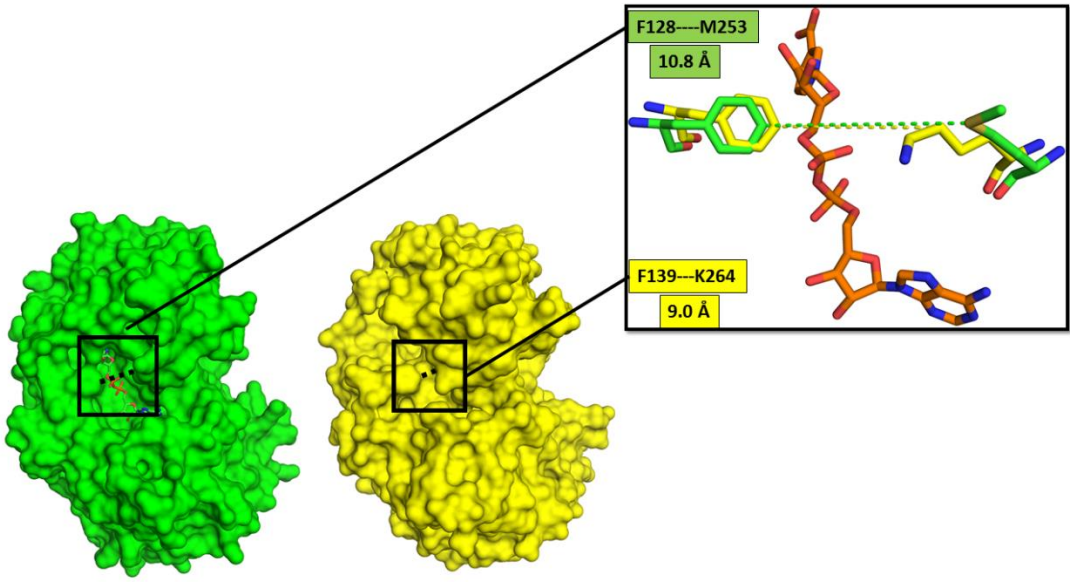


Figure 4.20: SpNadE NaAD binding site size comparison. The comparison of spNadE (yellow) and bsNadE (green) (PDB code: 2NSY: RMSD 0.85 Å) NaAD binding site residues. Distances are measured from the ζ carbon of F128 to the δ sulfur of M253 (bsNadE) and the ζ carbon of F139 to the δ carbon of K264. Structure comparisons indicate that when NaAD is bound the site is approximately 2.0 Å wider than when it is unbound.

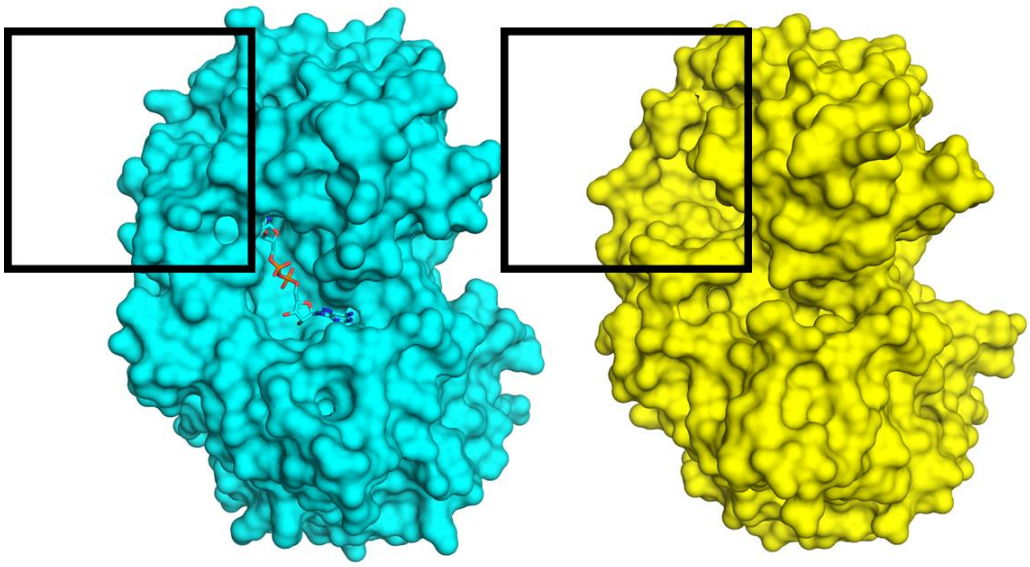


Figure 4.21: SpNadE ATP active site comparison. Cartoon representation spNadE and bsNadE (cyan) (PDB code: 2NSY) with boxes showing how open the ATP binding site is when no substrate is bound. This observation lead to the assumption that ATP binds first in the formation of NAD^+, then the NaAD site opens after ATP binding (details shown in next figure).

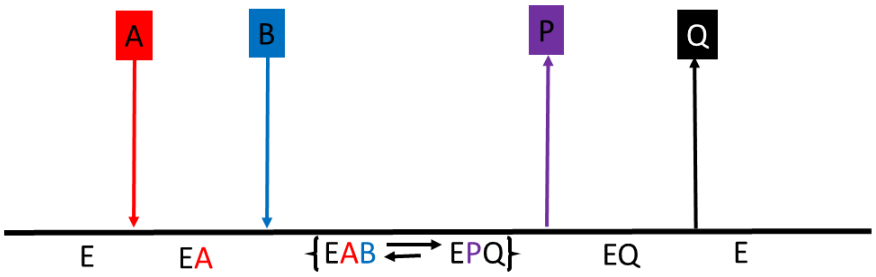


Figure 4.22: Cleland diagram of proposed general mechanism for spNadE. Schematic of a proposed ordered-sequential mechanism of enzymatic activity for spNadE from structural studies (20).

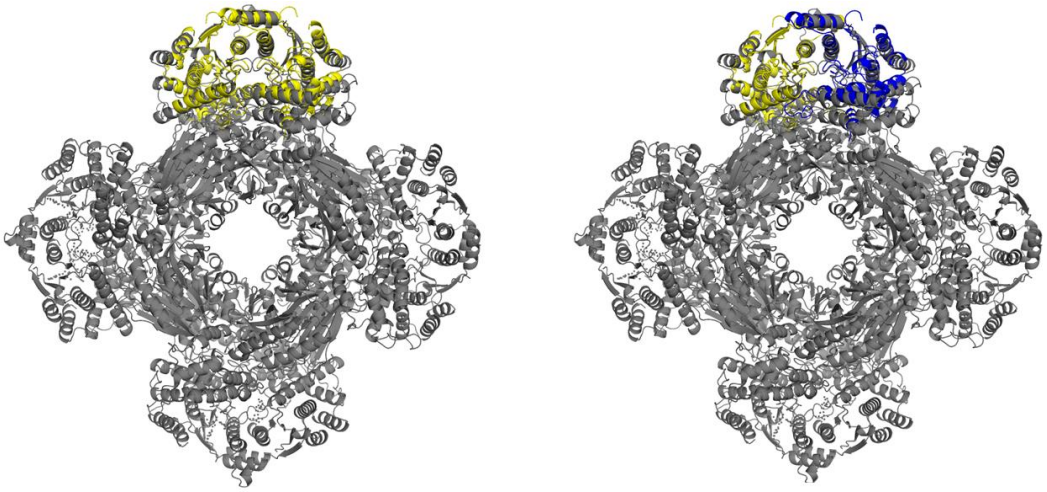


Figure 4.23: Comparison of spNade and mtNadE. Cartoon representation of the octameric, glutamine-dependent NAD^+ synthetase from *M. tuberculosis* (PDB code: 3SYT). On the left SpNadE (yellow) has been superposed on this structure to show its similarity to the NH_3 synthetase region of mtNadE. On the right, spNadE is colored by chain and this color schema is used in the following figures.

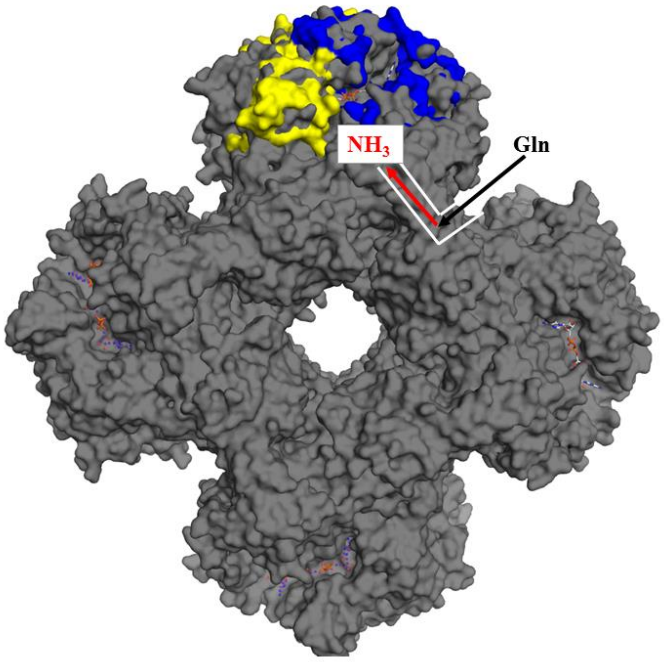


Figure 4.24: The glutamine and NH_3 channel for mtNadE. Surface representation showing superposition of of mtNadE (PDB code: 3SYT, grey) and spNadE (yellow and blue). This image provides a generalized diagram of the direction of the "tunnel system". The system is composed of two tunnels: the first binds Gln at the Glutaminase/ amidotransferase region and promotes the removal of NH_3, the second transfers NH_3 to the synthetase region for NAD^+ biosynthesis (9). Total length of the tunnel is approximately 73 Å (9).

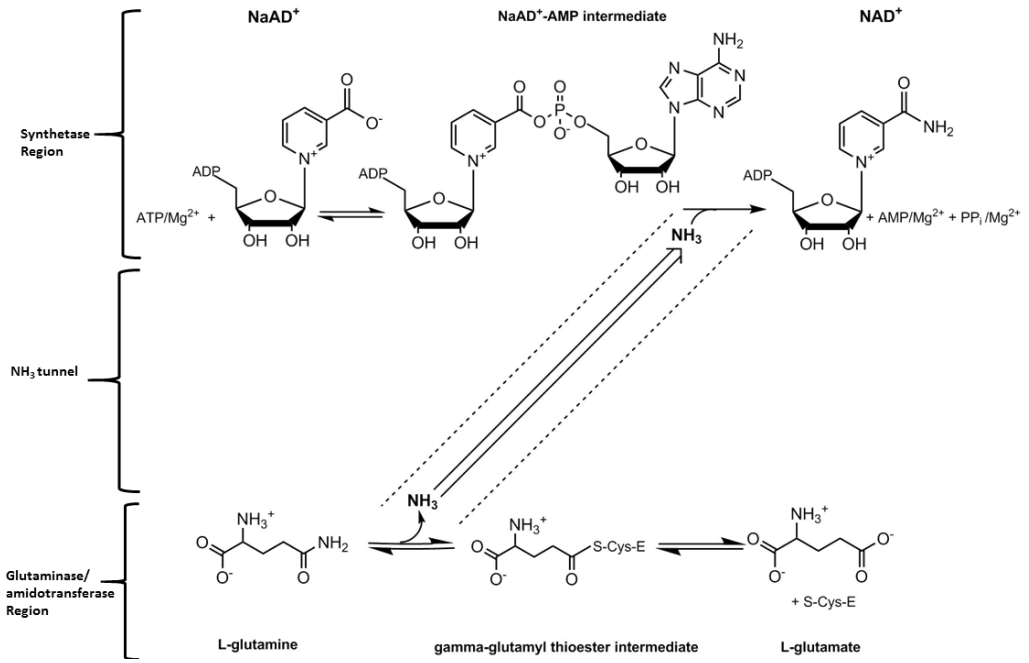


Figure 4.25: Schematic of the the NH_3 tunnel of mtNadE. The diagram adapted from LeRonde-LeBlanc *et al.* shows a general mechanism for NAD^+ biosynthesis in glutamine-dependent NAD^+ synthetases (9).

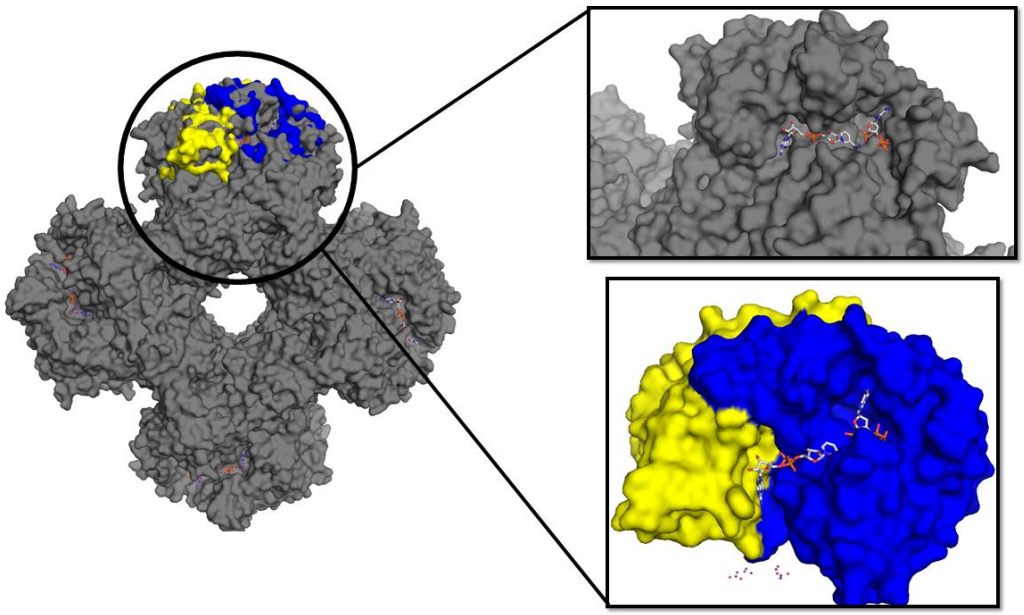


Figure 4.26: The proposed NaAD binding site in the mtNadE synthetase domain. SpNadE (yellow and blue) is superposed on mtNadE (PDB code: 3SYT; grey). NaAD is shown in stick representation. NaAD in spNadE was modeled using the mtNadE structure.

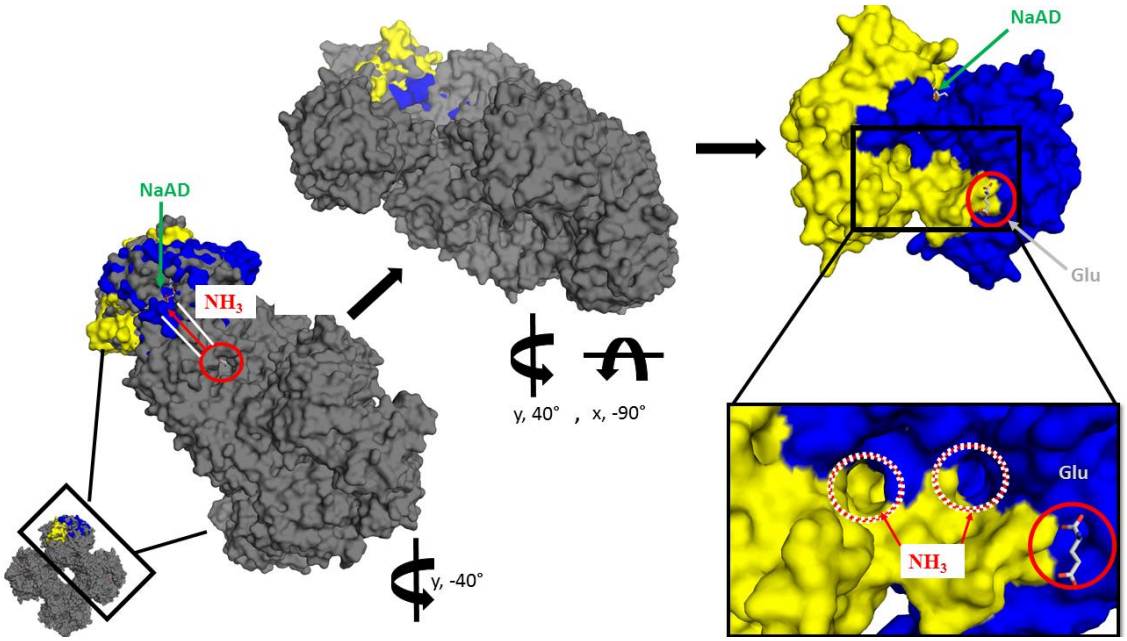


Figure 4.27: The discovery of the possible direction of NH_3 entrance in spNadE. Surface representation of spNadE superposed on mtNadE (PDB code: 3SYT) to show the proposed direction of travel of NH_3 from the glutaminase/ amidotransferase region to the NAD^+ synthetase region. The yellow/blue structure to the far right removes the superposed mtNadE structure to show evidence of two openings at the dimer interface of NadE that may provide insight into the direction of nucleophilic attack on the NaAD-adenylate intermediate that would be directly behind that interface during NAD^+ biosynthesis.

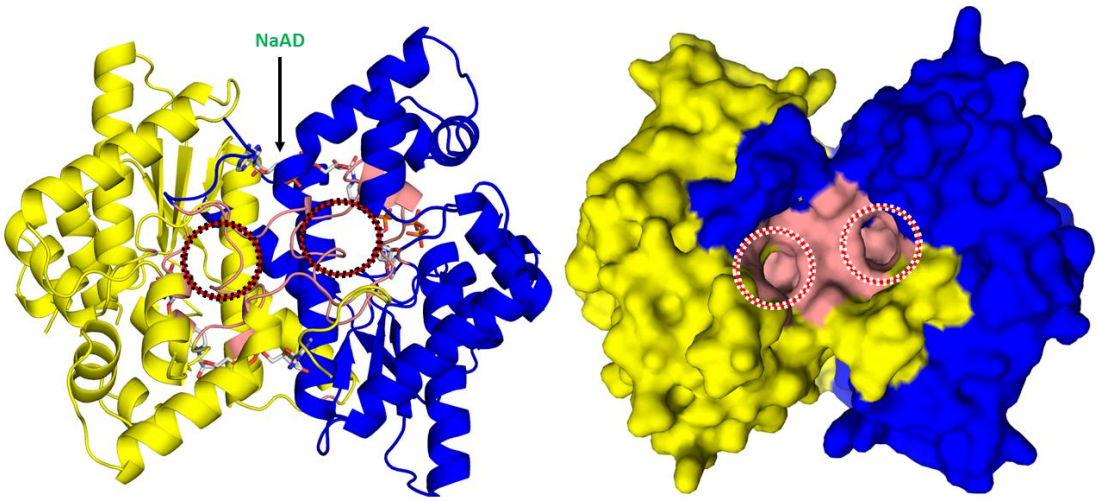


Figure 4.28: The proposed NH_3 passageway. Cartoon (*left)* and surface representations (*right*) of spNadE and NaAD molecules superposed from mtNadE (PDB code: 3SYT). The NaAD substrates are located behind the NH_3 passageway (highlighted in pink). Cavities, indicated by the circles, could explain the direction of nucleophilic attack of NH_3. We presume that NH_3 enters through one of these cavities in order to promote the conversion of NaAD to NAD^+.

4.5 REFERENCES

1. Sorci, L., I. K. Blaby, I. A. Rodionova, J. De Ingeniis, S. Tkachenko, V. De Crecy-Lagard, and A. L. Osterman. "Quinolinate Salvage and Insights for Targeting NAD Biosynthesis in Group A Streptococci." *Journal of Bacteriology* 195.4 (2012): 726-32. Web.
2. Mcdonald, Heather M., Pamela S. Pruett, Champion Deivanayagam, Irina I. Protasevich, W. Michael Carson, Lawrence J. Delucas, Wayne J. Brouillette, and Christie G. Brouillette. "Structural Adaptation of an Interacting Non-native C-terminal Helical Extension Revealed in the Crystal Structure of NAD Synthetase from Bacillus Anthracis." *Acta Crystallogr D Biol Cryst Acta Cryst D Acta Crystallogr Sect D Biol Crystallogr Acta Crystallogr Sect D Acta Crystallogr D Biol Crystallogr Acta Crystallographica Section D Biological Crystallography Acta Cryst Sect D Acta Crystallogr D* 63.8 (2007): 891-905. Web.
3. Symersky, Jindrich, Yancho Devedjiev, Karen Moore, Christie Brouillette, and Larry Delucas. "NH 3 -dependent NAD Synthetase from Bacillus Subtilis at 1 Å Resolution." *Acta Crystallogr D Biol Cryst Acta Cryst D Acta Crystallogr Sect D Biol Crystallogr Acta Crystallogr Sect D Acta Crystallogr D Biol Crystallogr Acta Crystallographica Section D Biological Crystallography Acta Cryst Sect D Acta Crystallogr D* 58.7 (2002): 1138-146. Web.
4. Jauch, R., A. Humm, R. Huber, and M. C. Wahl. "Structures of Escherichia Coli NAD Synthetase with Substrates and Products Reveal Mechanistic Rearrangements." *Journal of Biological Chemistry* 280.15 (2005): 15131-5140. Web.
5. Mao, F., P. Dam, J. Chou, V. Olman, and Y. Xu. "DOOR: A Database for Prokaryotic Operons." *Nucleic Acids Research* 37.Database (2009): n. pag. Web
6. Gough, Julian, Kevin Karplus, Richard Hughey, and Cyrus Chothia. "Assignment of Homology to Genome Sequences Using a Library of Hidden Markov Models That Represent All Proteins of Known Structure." *Journal of Molecular Biology* 313.4 (2001): 903-19. Web.
7. Baum, Leonard E., Ted Petrie, George Soules, and Norman Weiss. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains." *Ann. Math. Statist. The Annals of Mathematical Statistics* 41.1 (1970): 164-71. Web.
8. Bork, Peer, and Eugene V. Koonin. "A P-loop-like Motif in a Widespread ATP Pyrophosphatase Domain: Implications for the Evolution of Sequence Motifs and Enzyme Activity." *Proteins: Structure, Function, and Genetics Proteins* 20.4 (1994): 347-55. Web.
9. LaRonde-Leblanc, Nicole, Melissa Resto, and Barbara Gerratana. "Regulation of Active Site Coupling in Glutamine-dependent NAD Synthetase." *Nat Struct Mol Biol Nature Structural & Molecular Biology* 16.4 (2009): 421-29. Web.
10. Rodionova, I. A., B. M. Schuster, K. M. Guinn, L. Sorci, D. A. Scott, X. Li, I. Kheterpal, C. Shoen, M. Cynamon, C. Locher, E. J. Rubin, and A. L. Osterman. "Metabolic and Bactericidal Effects of Targeted Suppression of NadD and NadE Enzymes in Mycobacteria." *MBio* 5.1 (2014): n. pag. Web.
11. Luo, Hao, Yan Lin, Feng Gao, Chun-Ting Zhang, and Ren Zhang. "DEG 10, an Update of the Database of Essential Genes That Includes Both Protein-coding Genes and

Noncoding Genomic Elements: Table 1." *Nucleic Acids Research Nucl. Acids Res.* 42.D1 (2013): n. pag. Web.
12. Majorek, Karolina A., Misty L. Kuhn, Maksymilian Chruszcz, Wayne F. Anderson, and Wladek Minor. "Double Trouble-Buffer Selection and His-tag Presence May Be Responsible for Nonreproducibility of Biomedical Experiments." *Protein Science* 23.10 (2014): 1359-368. Web.
13. Qazi, S. Junaid S., Raymond Chew, Denice C. Bay, and Raymond J. Turner. "Structural and Functional Comparison of Hexahistidine Tagged and Untagged Forms of Small Multidrug Resistance Protein, EmrE." *Biochemistry and Biophysics Reports* 1 (2015): 22-32. Web
14. Nessi, C., A. M. Albertini, M. L. Speranza, and A. Galizzi. "The OutB Gene of Bacillus Subtilis Codes for NAD Synthetase." *Journal of Biological Chemistry* 270.11 (1995): 6181-185. Web.
15. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. Mcwilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. "Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology* 7.1 (2014): 539. Web.
16. Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, Z. Zheng, and S. H. Bryant. "CDD: A Conserved Domain Database for Protein Classification." *Nucleic Acids Research* 43.Database Issue (2015): 222-26. Web.
17. Robert, X., and P. Gouet. "Deciphering Key Features in Protein Structures with the New ENDscript Server." *Nucleic Acids Research* 42.W1 (2014): n. pag. Web.
18. Altschul, S., T. L. Madden, A. A. Schaffer, J. Zhang, W. Miller, and D. J. Lipman. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25.17 (1997): 3389-402. Web.
19. Krissinel, Evgeny, and Kim Henrick. "Detection of Protein Assemblies in Crystals." *Lecture Notes in Computer Science Computational Life Sciences* (2005): 163-74. Web.
20. Devedjiev, Yancho, Jindrich Symersky, Raj Singh, Marek Jedrzejas, Christie Brouillette, Wayne Brouillette, Donald Muccio, Debasish Chattopadhyay, and Larry Delucas. "Stabilization of Active-site Loops in NH 3 -dependent NAD Synthetase from Bacillus Subtilis." *Acta Crystallogr D Biol Cryst Acta Cryst D Acta Crystallogr Sect D Biol Crystallogr Acta Crystallogr Sect D Acta Crystallogr D Biol Crystallogr Acta Crystallographica Section D Biological Crystallography Acta Cryst Sect D Acta Crystallogr D* 57.6 (2001): 806-12. Web.
21. Cook, Paul F., and W. W. Cleland. *Enzyme Kinetics and Mechanism*. London: Garland Science, 2007. Print.
22. Cantoni, Rita, Manuela Branzoni, Monica Labo, Menico Rizzi, and Giovanna Riccardi. "The MTCY428.08 Gene of Mycobacterium Tuberculosis Codes for NAD1 Synthetase." *Journal of Bacteriology*, 180.12 (1998): 3218-221. Print.
23. Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. "Features and Development of Coot." *Acta Crystallogr D Biol Cryst Acta Cryst D Acta Crystallogr Sect D Biol Crystallogr Acta Crystallogr Sect D Acta Crystallogr D Biol Crystallogr Acta*

Crystallographica Section D Biological Crystallography Acta Cryst Sect D Acta Crystallogr D 66.4 (2010): 486-501. Web.

CHAPTER 5

THE STRUCTURAL CHARACTERIZATION OF SAGB, SAGC, SAGD AND SAGG PROTEINS FROM *STREPTOCOCCUS PYOGENES*

5.1 INTRODUCTION

In GAS the proposed SagBCD complex, composed of proteins expressed from the *sag* operon: SagB, SagC, and SagD, is considered to be primarily responsible for the biosynthesis of SLS (1, 2). More specifically, as suggested by Molloy *et al.*, the expression of the *sagA* reading frame, of the *sag operon,* results in the translation of a 53 amino acid strand (SagA). The strand is then converted into Streptolysin S (SLS) by the activity of the SagBCD. As determined through sequence homology studies, the SagA strand is divided into two segments: the leader peptide segment (residues M1-G23) and the protoxin segment (residues C24-T53). The leader segment contains an FXXXB motif (F4-I8: FTSNI) that is considered to be a recognition sequence for truncation of the strand by SagE (1, 3) (Figure 5.1). The mechanism for this truncation event has yet to be confirmed, but is presumed to be one of two events that is necessary in the conversion of SagA into SLS (1, 2). The second event involves the modification of cysteine, serine, and threonine residues within the protoxin segment (2). The chemical mechanism for this event is also undefined. The protoxin sequence (…↓CCCCCTTCCFSIATGSGNSQGGSGSYT) is considered to continue thru biosynthetic modification by action of SagBCD (Figure 5.2). Within the complex, SagC, a cyclodehydratase, is considered to act first in the conversion of the

cysteine, serine, and threonine residues into thiozole, oxazole, and metyloxazole rings, respectively (2). SagB is a proposed dehydrogenase, using flavin mononucleotide (FMN) as a cofactor, and is responsible for double-bond formation, within the rings (Figure 5.2). This step is considered the "finishing step" in the conversion of SagA into SLS. SagD is considered to be the docking protein that keeps the complex together to allow for SLS biosynthesis (1, 2) Sequence homology studies have identified the possibility of ATP being required for the activity of SagD (4). This interaction has yet to be confirmed.

Research interests also lie in the determination of the structure and function of SagG. SagG is a membrane associated protein proposed to play a role in the formation of the, membrane-bound, ATP-binding cassette. This cassette is considered responsible for the release of SLS into the extracellular environment (1, 2).

Currently the Protein Data Bank (PDB) contains no other structures with significant sequence similarity (over the entire sequence) to SagB, SagC, SagD, or SagG (Refer to Chapter 1, Table 1.2)

This chapter will provide a description of production of SagB, SagC, SagD and SagG proteins, as well as approaches used in an attempt to obtain their crystals.

5.2 Results and Discussion

Protein Production

SagB-D plasmid construction, protein expression, and purification protocols were the same as described for QSEs (Chapter 2, pages 27-30). In contrast to QSE isolation results, isolation attempts produced only microgram quantities of soluble protein, per liter

of culture (Figures 5.3- 5.5). SagD gel filtration results provided evidence of possible dimer formation, when compared to peaks from protein standards (results not shown). With the assistance of an undergraduate volunteer, David Mysona, we were able to isolate, at least, 5 mg of SagG per liter for culture (Figure 5.6). Additionally, SagG was found to be unstable in the imidazole-containing elution buffers, so immediate dialysis was required upon elution. Attempts to modify the growth conditions by changing IPTG concentrations, growing cultures in minimal media, decreasing inoculation temperature, and over-expression of the protein in different cell lines were used in attempts to improve folding and stability. None of the attempts were successful at improving protein yield in any of the Sag proteins.

Attempts were also made to isolate homologous proteins from different bacteria and different GAS strains. Work was initiated for the overexpression of the bacteriocin biosynthesis cyclodehydratase protein from *Staphylococcus aureus* in BL21 (DE3) *E. coli*. This protein is a structural homolog of SagC is predicted to be a more stable protein construct for crystallization (5). To date, attempts to isolate this protein have been unsuccessful.

A SagD analog was, also, identified from the M6 strain from GAS from the DNASU plasmid repository (Tempe, AZ). This construct is referred to as "SagD_DNASU". The plasmid is designed to express for Kanamycin-resistance and a 6x, C-terminal, polyhisitidine tag. There is a variation in molecular weight between SagD, at 51 kDa, and SagD_DNASU, at 45 kDa. More specifically this difference is due to a 50 amino acid truncation at the N-terminus of the sequence and a 48 residue extension at the C-terminus (residues 346-394) of SagD_DNASU. In addition, the protein from DNASU

carries a tyrosine mutation at residue 124 (Figure 5.6). Isolation attempts produced approximately 24 mg of protein per liter of culture, (after dialysis). Unfortunately, as determined from gel filtration results, a large percentage (~90%) of the isolated product was considered aggregate protein. A non-aggregated fraction (not shown) that was confirmed to contain SagD_DNASU and was screened for crystal formation, but no crystals have been obtained to date.

Refolding attempts on all of the Sag proteins were conducted and provided marginal success. The refolding conditions were partially based on Thermo Scientific Pierce® Protein Refolding Kit, and the conditions are listed in Table 5.1 (14). The most optimal conditions for refolding, for SagB, were solutions 2 and 3. For SagC, no precipitation was observed in solution 9, and for SagD solutions: 1, 2, 3, 4, 6, 8, 14, 15 were most optimal for refolding. Unfortunately, when gel filtration/ size exclusion chromatography was conducted, the peaks were presumed to be evidence of non-specific aggregation.

SagG was most stable in refolding condition 12. After the elution from the Ni-NTA column the protein was approximately 85-90% pure. As a result, the protein was further purified using gel filtration. The refolding of SagG proved to be only moderately successful as we were able to isolated only ~1 mg of pure protein from an, initial, 1L culture.

CRYSTALLIZATION

Due to protein aggregation and low protein yields only, approximately, 1000 combined crystal screening experiments were conducted for SagB-G.

MOLECULAR CLONING OF SAGBCD

The marginally successful attempts at SagB, SagC and SagD isolation prompted us to develop a methodology for expression of the entire SagBCD complex. Our approach was based on the strategy presented by Markley *et al.* (2012) (6). For the co-expression of the Sag proteins we enacted two slightly different approaches. Results for Strategy 1, SagBCD co-expression using two plasmids, showed that SagB was successfully introduced into the pETDuet™ (Novagen, Madison, WI) plasmid and it has been confirmed through sequencing (Figure 5.8). There were numerous attempts, to insert SagD into the second multi-cloning site, in the plasmid, but attempts have been unsuccessful. However there was success with the addition of new restriction sites to the SagD insert (Figure 5.9). Results for Strategy 2, SagBCD expression from 3 individual plasmids, showed some success as evidenced by growth of transformed BL21 (DE3) *E. coli* on the Ampicillin/Kanamycin/Chloramphenicol infused LB agarose plates. The introduction of each of the individual plasmids still needs to be confirmed through sequencing (refer to Figure 5.10 for agarose gel results). Culture and over-expression of the proteins showed evidence of protein production (not shown) although the product was completely insoluble.

HOMOLOGY MODELING

We used PHYRE2 for homology modeling and the approach described in the homology modeling section in Chapter 3, page 60. Collectively, structure confidence parameters for top homology models was greater than ≥99 %, but sequence identity for models ranged from 10- 41%. The homology models, with quality parameters of 100% confidence, included templates from: *Thermotoga maritima* (PDB code: 4EO3; sequence identity: 18%; 69% sequence coverage) for SagB,(Figure 5.11), *Escherichia coli* (PDB code: 3H9G; sequence identity: 10%; 91% sequence coverage) for SagC (Figure 5.12),

Prochloron sp. 06037A (PDB code: 4BS9; sequence identity 23%; 79% sequence coverage) for SagD (Figure 5.13), and *Thermotoga maritima* (PDB code: 4YER; sequence identity 41%; 74% sequence coverage) for SagG (Figure 5.14). The results of the homology modeling suggest that SagG, SagB, SagC and SagD proteins most likely dimerize. Therefore, one may speculate that the enzymatic complex responsible for SLS formation has composition $(SagB)_2(SagC)_2(SagD)_2$ and not SagBCD as reported in the literature.

FUTURE DIRECTIONS

In order to increase the likelihood of Sag protein isolation and overcome issues with solubility new strategies must continue to be initiated. New approaches include: introducing substrates (i.e. ATP and/or SagA) into the culture media during growth and over-expression. The assumption is that through substrate interaction the proteins can be stabilized and could allow for higher yields of soluble during isolation. SagD was identified to be a similar to proteins that are a part of the YcaO family of proteins which are considered responsible for the formation of thiazole and oxazole modifications in bacteria; they are proposed to require ATP for activity (4, 7).

To improve protein yield utilizing conditions that can slow the rate of protein over-expression could be conducive to improving solubility. Possible avenues of approach could be the growth of the transformed bacteria in minimal media and/or the use of Arctic Express Cells (Agilent Technologies, Santa Clara, CA) which contain chaperones proteins that can allow for folding at temperaturs ranging between 10- 13 °C.

Another strategy involves protein engineering and the addition of maltose-binding protein (MBP) to the N or C-terminus of each of the Sag proteins. Maltose binding proteins has been shown to improve solubility of unstable proteins, when attached (8). Deletions

of predicted unstructured fragments and/ or a change of the His-tag location may also be useful approaches. (9, 10, 11)

While the expression of all three of the proposed SagBCD complex proteins have been completed, in one cell, continued work will need to be done to improve yield of the complex proteins in order to visualize interactions, if any, through x-ray crystallography. Continued work will also be done to attempt to introduce SagD into the pET-Duet plasmid that contains SagB. Introduction of new restriction sites have been completed for SagD (Figure 5.7), but a more optimized strategy for ligation needs to be prepared to successfully insert the SagD into pET-Duet+SagB for co-expression. Strategies to overcome this could include the use of temperature-cycle ligation strategy created by Lund *et al.* (12).

5.3 Materials and Methods

Protein Production and Crystallization

All experiments for transformation, overexpression, isolation and crystallization utilized the strategies outlined in Chapter 1, pages 27-30.

Protein Refolding and Isolation

After initial attempts at isolation proved unsuccessful at producing a significant amount of protein for crystal screening. The a small portion of the insoluble fraction (~10 mg) was dissolved in 6 M Guanidine HCl or 8 M Urea and each of the Sag proteins were re-isolated by IMAC in denaturing conditions.

After elution of the unfolded protein from the column, approximately 100 µL of eluate, in denaturing conditions, was added to 10 mL of the re-folding conditions reported in Table 5.1. This approach is called fast dilution re-folding and was reported to be

successful for some proteins (13, 14). After the dilution the solution of the protein was stirred overnight (~12 hrs) at room temperature. The refolding condition that produced the least of protein precipitate were selected for further experiments. The solution(s) identified during these experiments were used to store the protein and perform gel filtration purification. After gel filtration, the protein was checked for purity using SDS-PAGE, concentrated and used for crystallization experiments.

MOLECULAR CLONING OF SAGBCD COMPLEX

Due to the instability of the Sag proteins, it was assumed that through co-expression of the proteins, there could be an increase in the stability of the individual proteins as a result of possible stabilizing interactions from complex formation. To accomplish this, two strategies were enacted; in Strategy 1 the goal was to introduce *sagB*, *sagD* into a single plasmids and allow it to be co-expressed with *sagC*. More specifically, one of the plasmids would be engineered to co-express SagB and SagD in a plasmid encoding for Ampicillin-resistance (the pETDuet™ plasmid), while SagC would be expressed on a separate plasmid encoded to promote Kanamycin resistance (Figure 5.15). In Strategy 2, the intention would be to co-express each plasmid individually, within the same organism. Each plasmid would contain genetic information for different antibiotic resistance, more specifically, SagB would be introduced into a plasmid encoding for Ampicillin-resistance, SagC into a plasmid encoding for Kanamycin-resistance, and SagD into a plasmid encoding for Chloramphenicol resistance (Figure 5.16).

These protocols were adapted from published strategies from Tolia and Joshua-Tor (2006) and Markley *et al.* (2011) and are explained in greater detail below (6, 15).

Strategy 1:

Plasmids for SagB, SagD, and pET-DUET (Novagen, Madison, WI) were transformed into competent DH5 alpha cells (NEB, Ipswich, MA) using the heat/ cold shock technique (16). A single colony from each transformation experiment was cultured in 5 mL of LB media, with respective antibiotic, for 12-16 hours. Each culture was centrifuged to at 3900 *g* for 10 min and the culture media was discarded. Plasmids were extracted and concentrated according to the protocol for the GeneJET plasmid Miniprep kit (Thermo Scientific, Waltman, MA). Plasmid concentrations were determined by UV spectroscopy, and eventually stored at -20 ºC for future testing.

For the introduction of SagB into the pETDuet™ plasmid, restriction enzymes NdeI and XhoI were identified for excision and experiments were conducted in accordance with protocol from NEB Double digest finder (17). At experiment completion, agarose gel electrophoresis was used to isolate cut pETDuet™plasmid and excised SagB. Upon identification of SagB and pETDuet™, they were cut out of the gel and made soluble through the use of the Plasmid GeneJET Gel extraction Kit (Thermo Scientific, Waltman MA). Ligation experiments were prepared in 6, 20 µL experiments. Experiment conditions all contained a standard, 2 µL of SagB DNA, 2 µL of buffer, and 1 µL of T4 DNA Ligase (NEB). Variation in the experiment preparation occurred with increasing volumes of plasmid DNA in the following consecutive volumes, 0.5 µL, 1 µL, 1.5 µL, 2 µL, 2.5 µL, and 3 µL, with accommodating amounts of nuclease-free water to achieve a 20 µL total volume. Ligation experiments were run overnight at 18 °C. Due to low recovery from the gel ligation, experiments went directly to the transformation protocol. Any colony growth that was obtained was cultured and confirmed by sequencing.

SagD, within the pjexpress411 plasmid (DNA 2.0, Menlo Park CA), did not contain useful restriction sites for introduction into the pETDuet™system, so new restriction sites had to be created. Primers for restriction sites EcoRI and NdeI created and were introduced by PCR using the following protocol: **Cycle # 1**: 98 °C: 2 min, **Cycle # 2-33**: Denature- 98 °C, Anneal-65 °C- 55 °C: 30 sec, Extend-72 °C:1 min, 30 sec, **Final Cycle (# 34)**: Denature-98 °C, Anneal-65 °C, Extend 72 °C, 5 min, Final Hold at 4 °C. Primer information is shown in Figure 5.17.

Strategy 2:

In strategy 2, SagB (Ampicillin-resistant) and SagC (Kanamycin-resistant) plasmids were introduced into BL21 (DE3) cells, simultaneously, using the heat shock/ cold shock method (16). Cells were cultured and grown on Amp/Kan resistant plates for confirmation. SagBC cells were then made competent using a calcium chloride method adapted from Sambrook *et al.* (18). SagD, which has Kanamycin resistance in the pJexpress plasmid, had to be transplanted into another antibiotic-resistant plasmid to confirm successful transformation into the new SagBC strain. The SagD insert was cut from the pJexpress41 plasmid (DNA 2.0, Menlo Park, CA) using the restriction enzymes XhoI and XbaI and transferred into a pHSG396 plasmid which contained genomic information for Chloramphenicol resistance. The transfer of the insert was conducted utilizing the same techniques outlined in Strategy 1. The introduction of SagD insertion into the Cam-res plasmid was confirmed by sequencing (Eton Biosciences, RTP, NC). The plasmid was introduced into the SagB/C competent cells, and introduction was confirmed by the growth of the bacteria on the Kan/Amp/Cam resistant plates.

OTHER COMPUTATIONAL METHODS

All Other Computational Methods are consistent with what was described in Chapter 2, page 33.

5.4 TABLES AND FIGURES

Table 5.1: Conditions used for protein re-folding (14).

Condition	Components
1	0.5 M L-arginine, 0.05 M MES (pH 6.0), 0.05 M NaCl, 0.05 M KCl
2	0.5 M L-arginine, 0.05 M Tris (pH 8.0), 0.05 M NaCl, 0.05 M KCl
3	0.5 M L-arginine (pH 8.5)
4	0.5 M guanidine HCl, 0.05 M MES (pH 6.0)
5	0.5 M guanidine HCl,0.05 M Tris (pH 8.0)
6	0.25 M KCl, 0.05 M HEPES (pH 6.0)
7	0.4 M sucrose, 0.05 M MES (pH 6.0)
8	0.4 M sucrose, 0.05 M Tris (pH 8.0)
9	1% w/v PEG 3350, 0.05 HEPES (pH 7.5), 0.01 M NaCl, 0.01 M KCl
10	1% w/v beta-cyclodextrin, 0.05 HEPES (pH 7.5), 0.01 M NaCl, 0.01 M KCl
11	0.5 M glycine ethyl ester HCl, 0.05 HEPES (pH 8.0), 0.01 M NaCl, 0.01 M
12	0.5 M NDSB-221, 0.05 HEPES (pH 8.0), 0.01 M NaCl, 0.01 M KCl
13	0.5 M NDSB-221, 0.05 HEPES (pH 8.0), 0.01 M NaCl, 0.01 M KCl
14	0.5 M L-arginine (pH 8.5) , 0.01 M $MgCl_2$, 0.01 M KCl
15	0.5 M L-arginine (pH 8.5), 0.01 M $CaCl_2$, 0.01 M KCl
16	0.5 M L-arginine (pH 8.5), 0.001 M EDTA, 0.01 M KCl

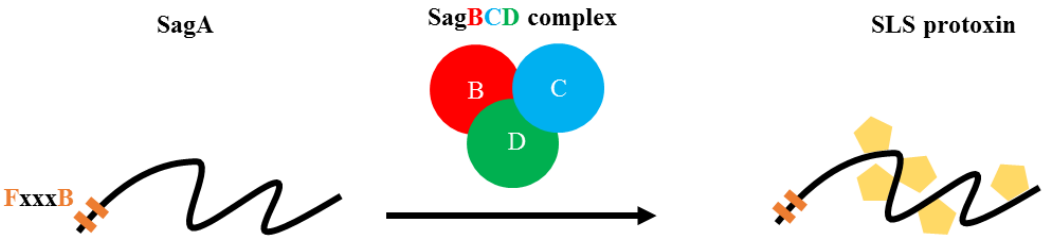


Figure 5.1: Overview of predicted SagBCD complex reaction. a schematic of the proposed activity of the proteins of the SagBCD complex (3). The FxxxB region of SagA is described as a possible recognition sequence by SagE (1). The FxxxB corresponds to phenylalanine followed by 3 residues and a basic amino acid residue B (3).

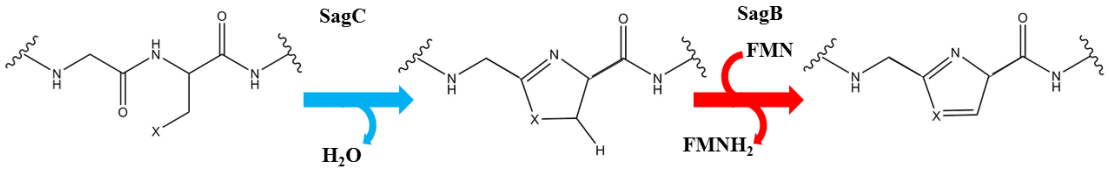


Figure 5.2: Schema of the proposed enzymatic activity of SagB and SagC. A schematic of the proposed activity of SagB and SagC (3). SagC is considered to operate as a cyclodehydratase forming the ring (3). SagB is suggested to function as a dehydrogenase, removing hydrogen from the ring with the assistance of flavin mononucleotie (FMN) to create reduced FMN ($FMNH_2$) (3).

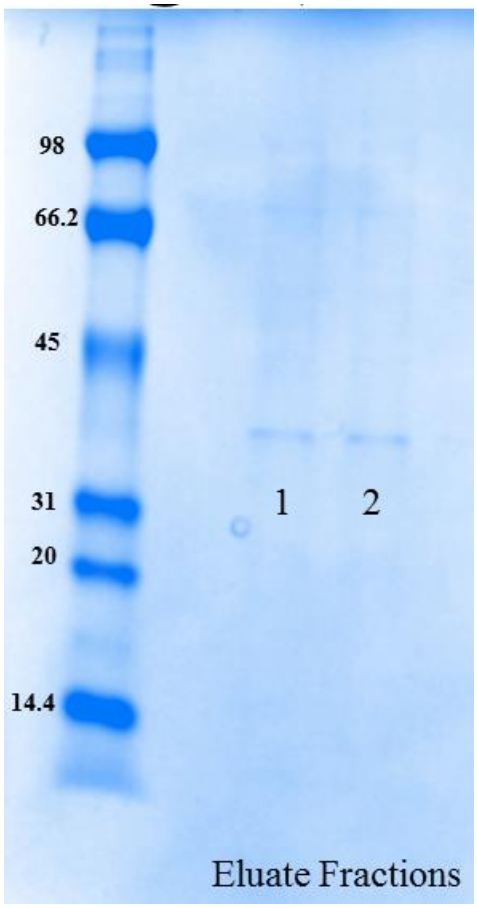


Figure 5.3: SagB SDS-PAGE isolation result. This image shows SDS-PAGE results confirming the isolation of SagB. Total volume obtained: ~ 1 mg/ mL. Sample volume applied to well: 10 µL. The molecular weights for marker poteins are reported in kDa.

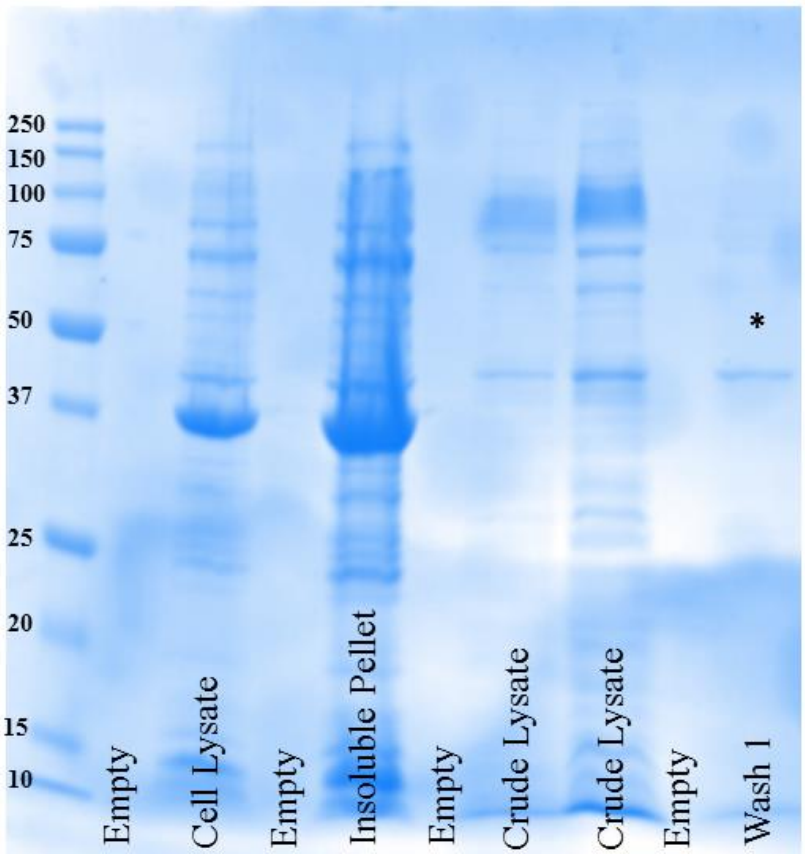


Figure 5.4: SagC SDS-PAGE isolation result. SDS-PAGE results confirming the isolation of SagC. Total volume obtained: ~ 1 mg/ mL. Sample volume applied to well: 10 µL. Isolated SagC is marked with an asterisk. The molecular weights for marker poteins are reported in kDa.

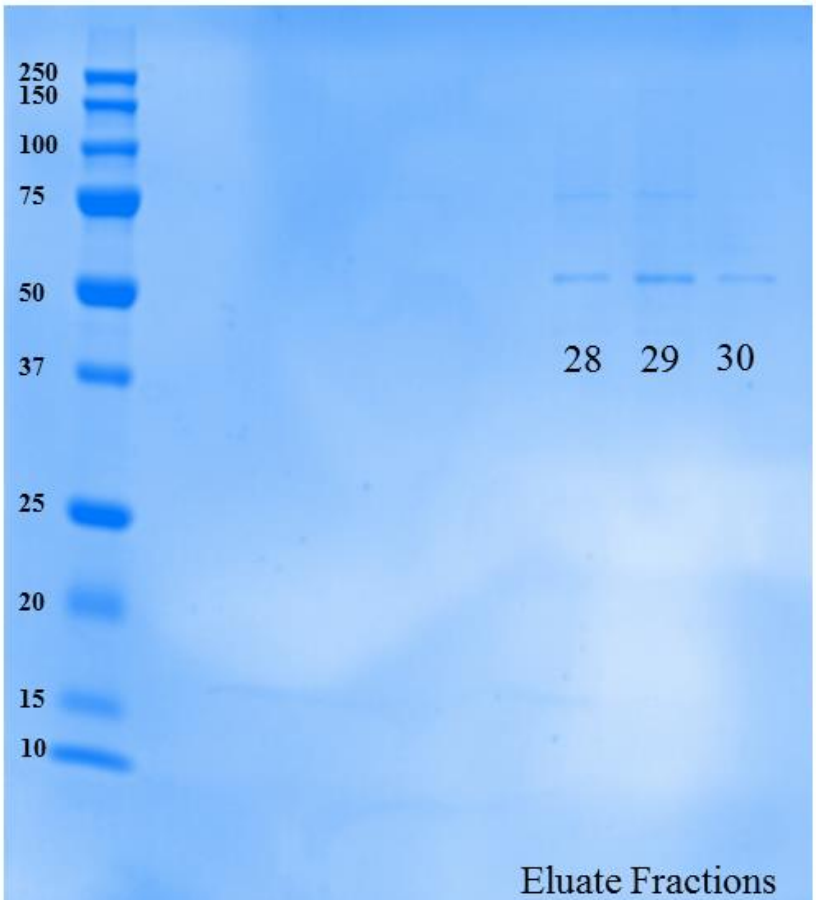


Figure 5.5: SagD SDS-PAGE isolation result. SDS-PAGE results confirming the isolation of SagD. Total volume obtained: 1 mL. Concentration: ~1 mg/mL. Sample volume applied to well: 10 µL. The molecular weights for marker poteins are reported in kDa.

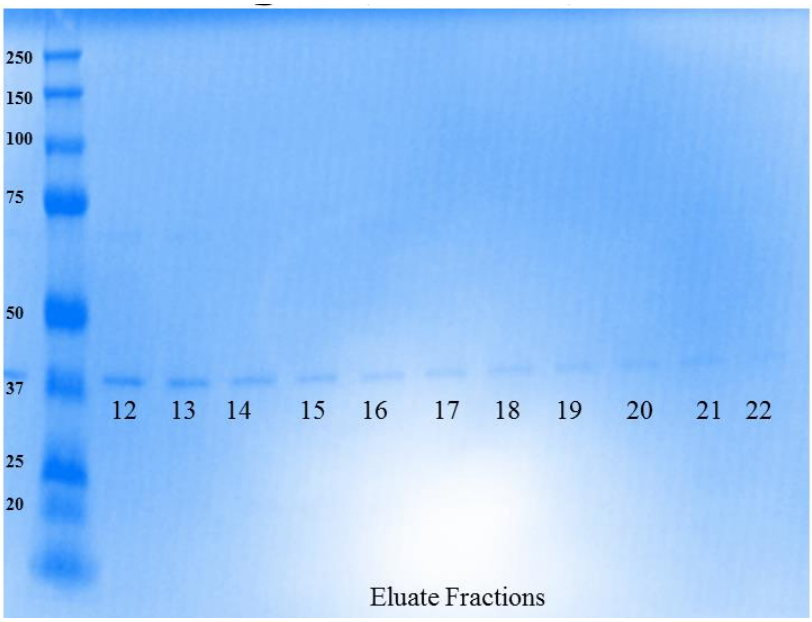


Figure 5.6: SagG SDS-PAGE isolation result. SDS-PAGE results confirming the isolation of SagG. Total volume obtained: 1 mL. Concentration: ~5 mg/mL. Sample volume applied to well: 10 µL. The molecular weights for marker poteins are reported in kDa.

```
SagD_DNA_2.0  1   MHHHHHHSSGVDLGTENLYFQSGSGMLYYYPSFNHIFDELKSLSGNRTGILNQSQVPVCN   60
SagD_DNASU    1   ------------------------------------------------------------    0

SagD_DNA_2.0  61  HPHDVYLKSITGQMPDYHKQFIGELSQVSYHIIGYGSYYEEALIKYLGESIERYATVIAG  120
SagD_DNASU    1   --------------DYHKQFIGELSQVSYHIIGYGSYYEEALIKYLGESIERYATVIAG   45
                              ******************************************

SagD_DNA_2.0  121 DLLSDRIVYASYNELKLLHKVMPLEYLQVFTQEQIALSCDLQMMMCDKMVTENDVLGWVK  180
SagD_DNASU    46  DLLSDRIVYASYNELKLLHKVMPLEYLQVFTQEQIALSCDLQMMMCDKMVTENDVLGWVK  105
                  ************************************************************

SagD_DNA_2.0  181 CPMFFEDAEMYVPAQMLCVGYKTNETVGERRIIPGFSTGTASHKTLEAAMCNSLIEYIQI  240
SagD_DNASU    106 CPMFFEDAEMYVPAQMLYVGYKTNETVGERRIIPGFSTGTASHKTLEAAMCNSLIEYIQI  165
                  ***************** ******************************************

SagD_DNA_2.0  241 DSMMLSWYTKKPCPKIIVDDPDIEVILEEARLGKDSLYDIIPIDMTVGEDNPLYTFGIIL  300
SagD_DNASU    166 DSMMLSWYTKKPCPKIIVDDPDIEVILEEARLGKDSLYDIIPIDMTVGEDNPLYTFGIIL  225
                  ************************************************************

SagD_DNA_2.0  301 KNKYDEGPYLLFGVQAGLDPKHALLRGIMEASAISYSYYYNLLYQKASLANIECEEPLFL  360
SagD_DNASU    226 KNKYDEGPYLLFGVQAGLDPKHALLRGIMEASAISYSYYYNLLYQKASLANIECEEPLFL  285
                  ************************************************************

SagD_DNA_2.0  361 DLDSNVFYYAHPKDQDHKWKAFEPLISGEVLLSDLEDHSGKDKKEDLKTLLAYAKKVSPN  420
SagD_DNASU    286 DLDSNVFYYAHPKDQDHKWKAFEPLISGEVLLSDLEDHSGKDKKEDLKTLLAYAKKVSPN  345
                  ************************************************************

SagD_DNA_2.0  421 ------------------------------------------------  420
SagD_DNASU    346 AVFLDITPPEALEKGWYVTRVLMPELLEMCIPAFPFANHPREGHHHHHH  394
```

Figure 5.7: SagD/SagD_DNASU sequence alignment results. Comparison of SagD (SagD_DNA_2.0) and SagD_DNASU sequences.

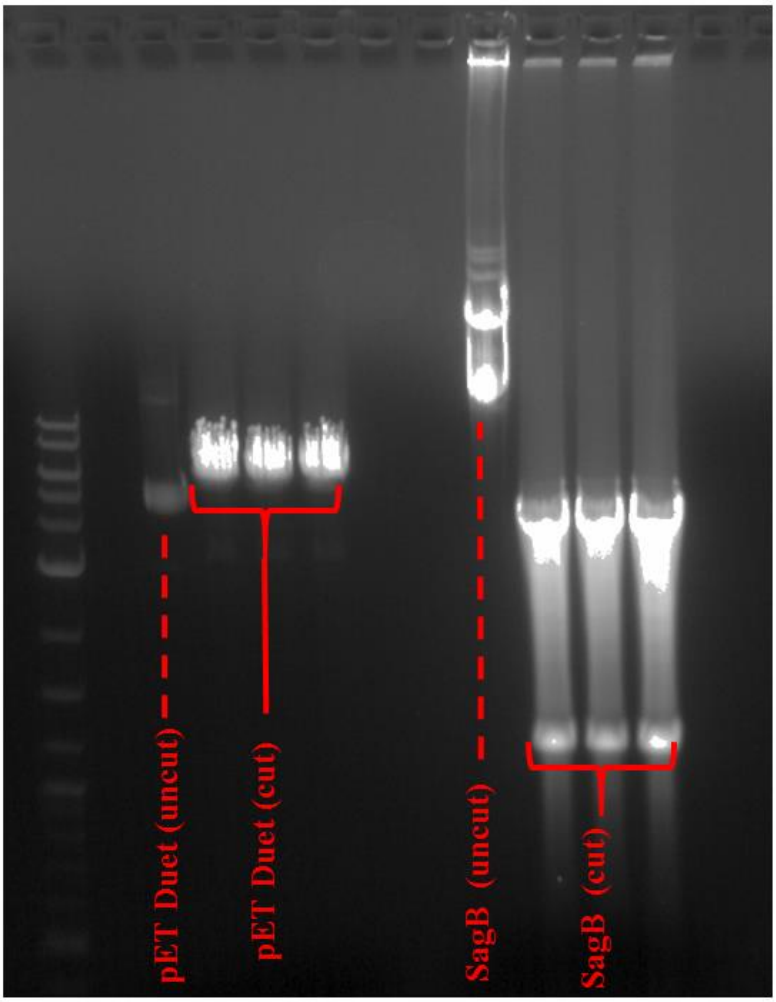


Figure 5.8: Agarose gel result for SagB and pETDuet™ restriction enzyme experiments (Strategy 1). This agarose gel result shows evidence of restriction enzyme activity on the pJexpress411 plasmid containing SagB plasmid, and the pETDuet™ plasmid. "SagB (cut)" was successfully ligated into the pETDuet™ plasmid.

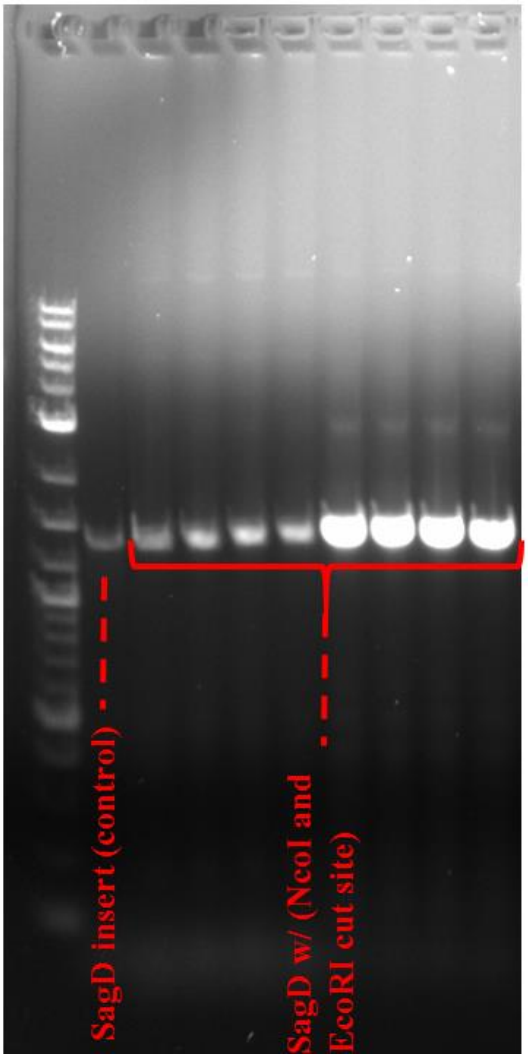


Figure 5.9: Agarose gel result for addition of new restriction sites for SagD (Strategy 1). The agarose gel shows evidence of amplification and isolation of the SagD insert with new EcoRI and NcoI restriction sites.

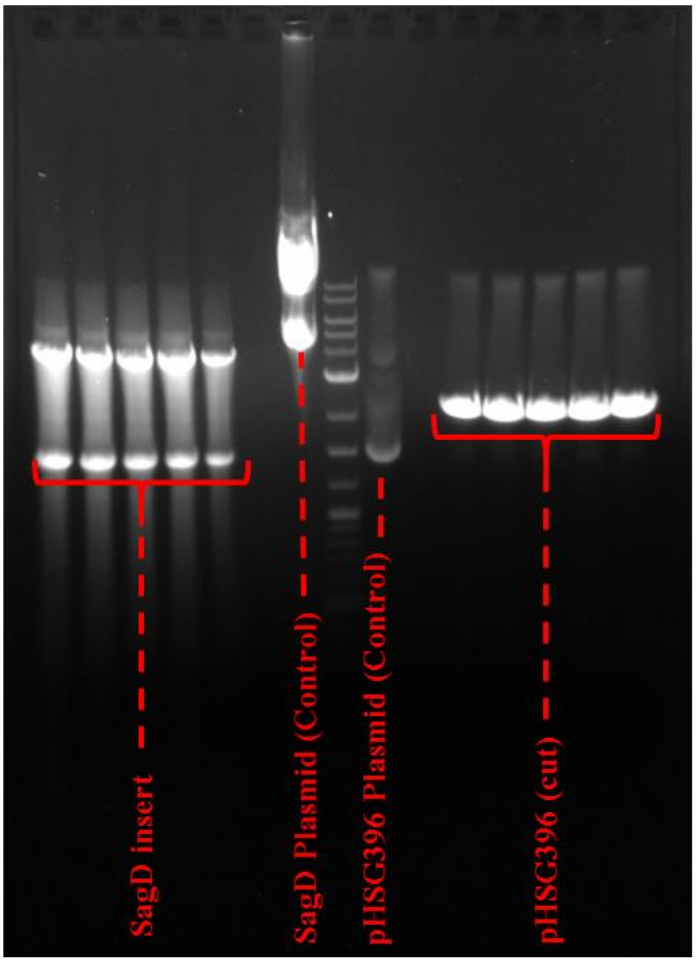


Figure 5.10: Agarose gel result for the SagD and pHSG396 restriction enzyme experiment (Strategy 2). The agarose gel presents the successful removal of the SagD insert from the pJexpress411 plasmid, as well as, results of the successful excision of the pHSG396 plasmid. The SagD insert was successfully transferred into the pHSG396 plasmid.

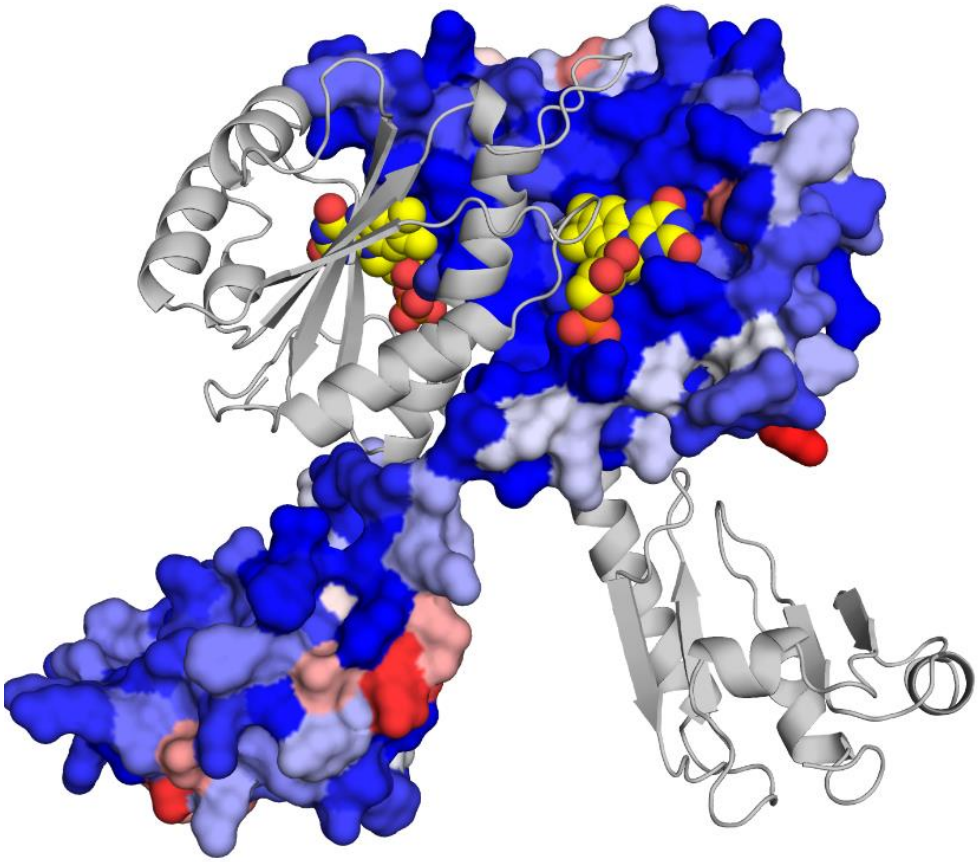


Figure 5.11: Homology modeling result for SagB. Calculated structure of SagB obtained from using PHYRE2. The homology modeling results indicate that SagB may form a dimer. The putative dimer is represented in cartoon and surface representations. The sequence conservation is mapped on the protein surface representation (blue color corresponds to highly conserved residues, while red indicates variable residues). Binding sites for FMN (yellow) are highly conserved in term of protein sequence. The structure is based on bacterioferritin comigratory protein from *Thermotoga maritima* (PDB code: 4EO3).

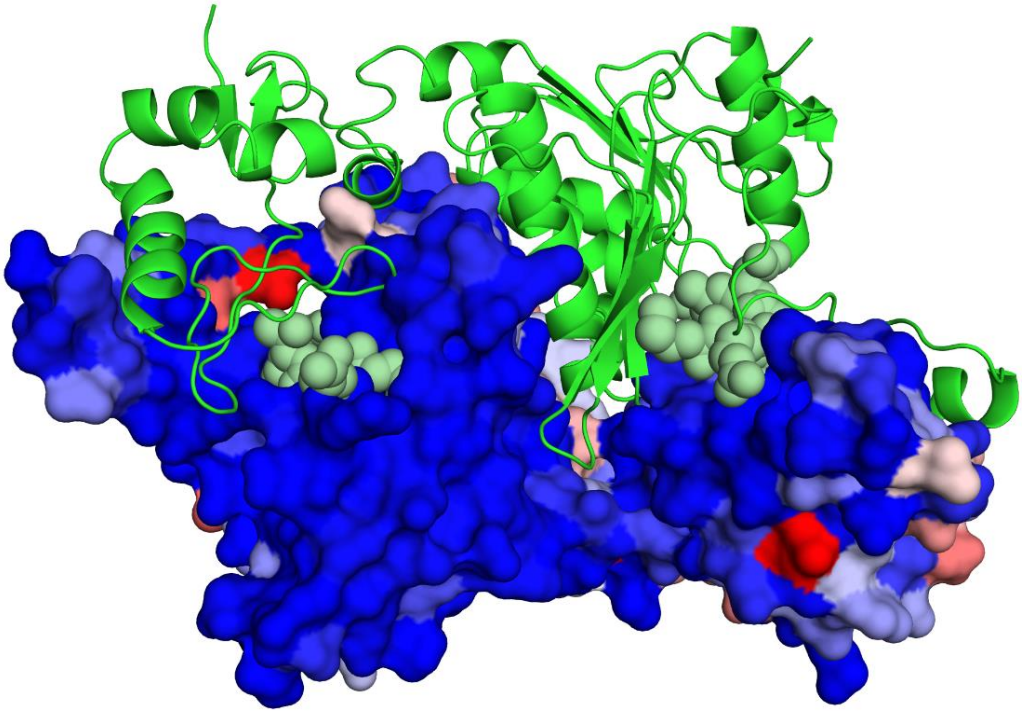


Figure 5.12: Homology modeling result for SagC. Calculated structure of SagC was obtained using PHYRE2. The modeling results also suggests that SagC may form a dimer. The putative dimer is represented in cartoon and surface representations. The sequence conservation is mapped on the protein surface representation (blue color corresponds to highly conserved residues, while red indicates variable residues). The structure is based on a complex between *E. coli* MccB protein and MccA-N7isoASN peptide (PDB code: 3H9G). The peptide (pale green) is a microcin C7 analog.

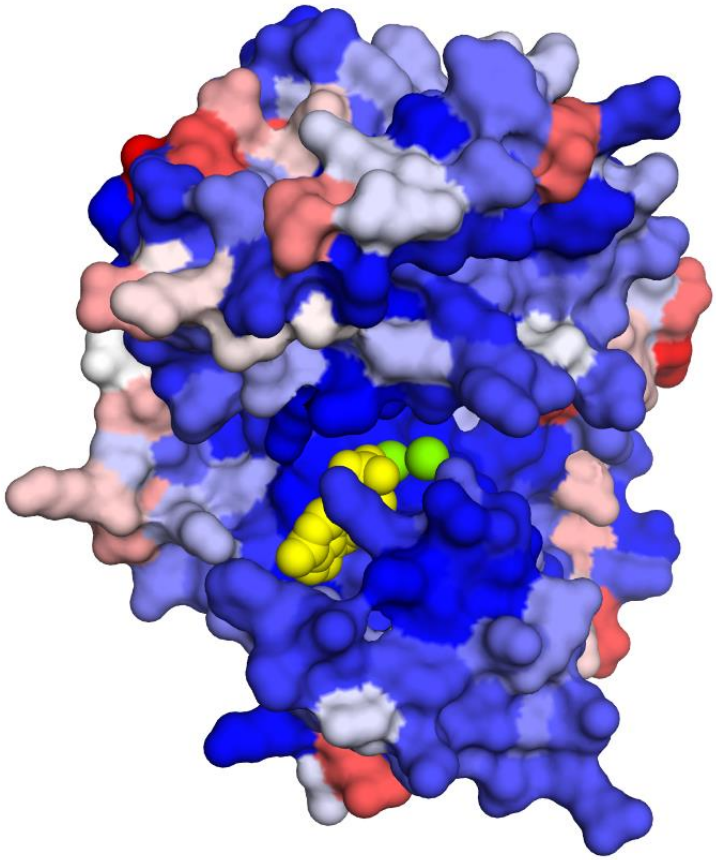


Figure 5.13: Homology modeling result for SagD. Calculated structure of SagD obtained using PHYRE2. The structure is based on structure of heterocyclase TruD from an uncultured *Prochloron sp*. 06037A (PDB code: 4BS9). Similar to the structures above, the homology model suggests that SagD may also form a dimer. The putative dimer is represented in cartoon and surface representations. The sequence conservation is mapped on the protein surface (blue color corresponds to highly conserved residues, while red indicates variable residues). AMP and Zn^{2+} (yellow-AMP / green-Zn^{2+}) positions were also modeled using the heterocyclase TruD structure.

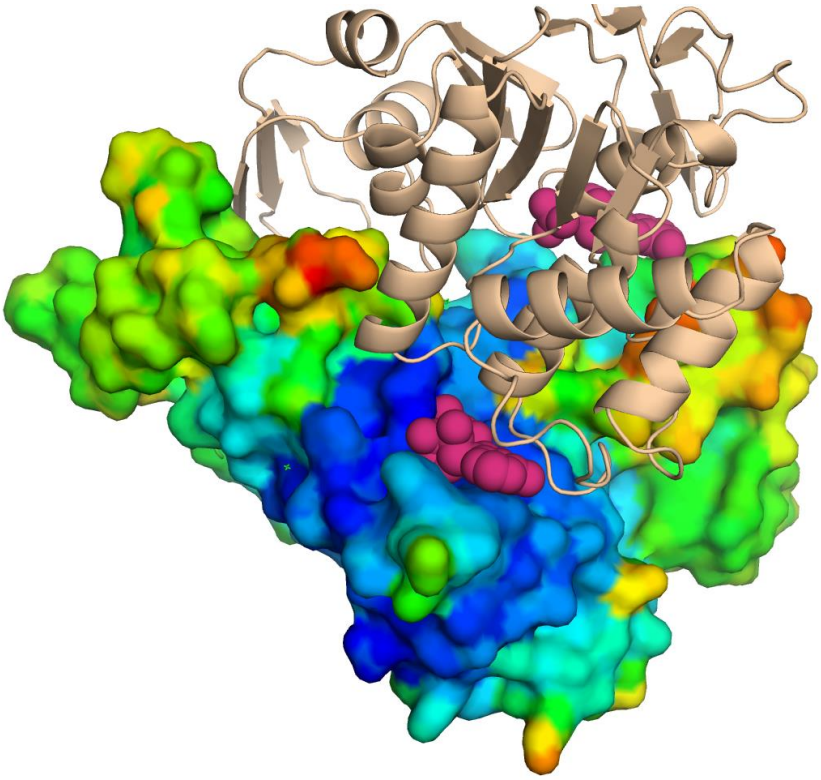


Figure 5.14: Homology modeling result for SagG. Calculated structure of SagG was obtained using PHYRE2. The structure is based on a structure of the ABC transporter ATP-binding protein from *Thermotoga maritima* (PDB code: 4YER). The homology model results predict that SagD is also a dimer with a proposed binding sites for ADP (pink). The putative dimer is represented in cartoon and surface representations. The sequence conservation is mapped on the protein surface representation (blue color corresponds to highly conserved residues, while red indicates variable residues).

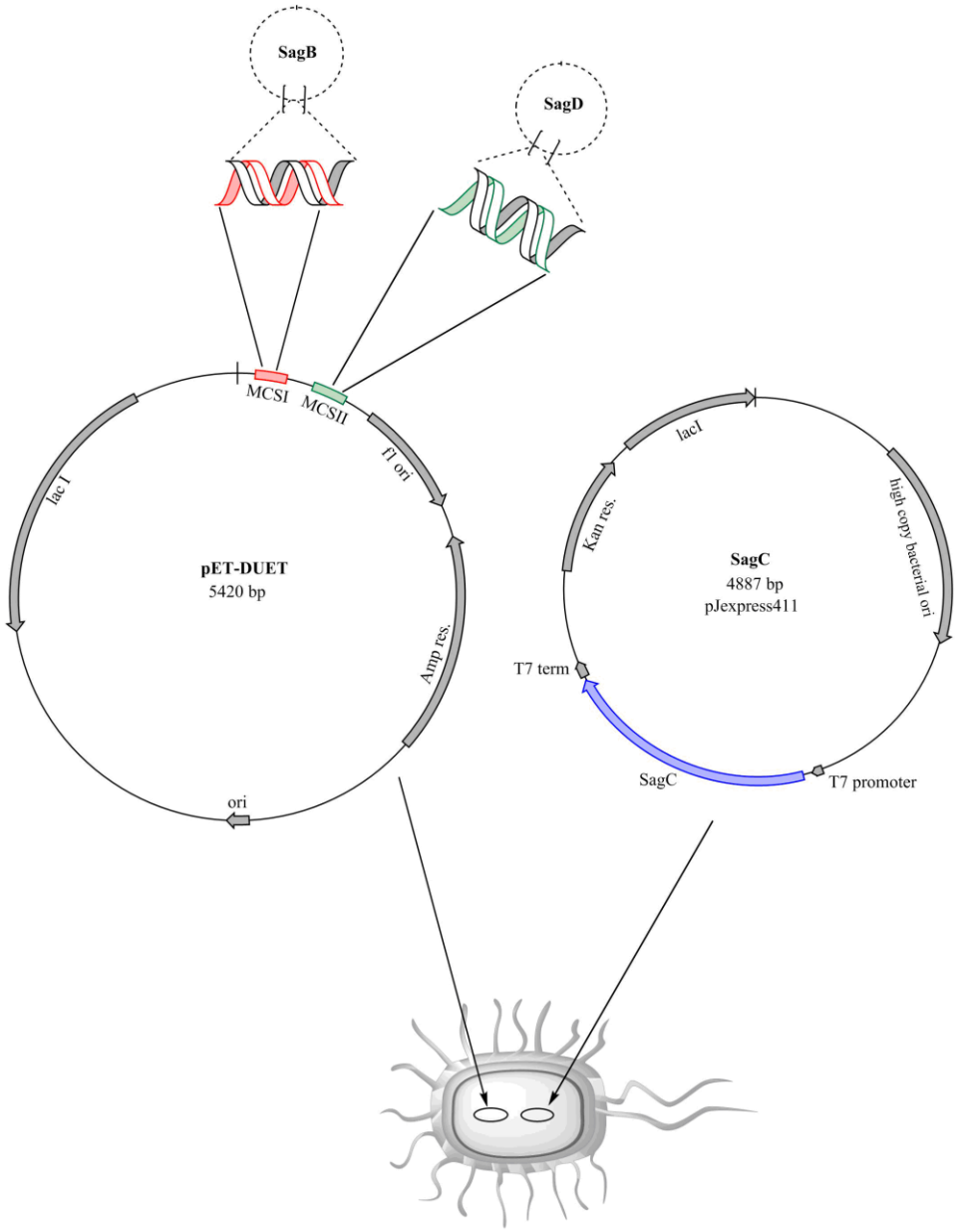


Figure 5.15: Schematic of SagBCD molecular cloning strategy 1. This is a diagram of Strategy 1 for SagBCD co-expression in *E. coli*. This strategy involves the insertion of two plasmids (one expressing for SagB and D, and the other for SagC only) into BL21 (DE3) *E. coli* for co-expression experiments. The experimental design was adapted from Markle *et al.* (11).

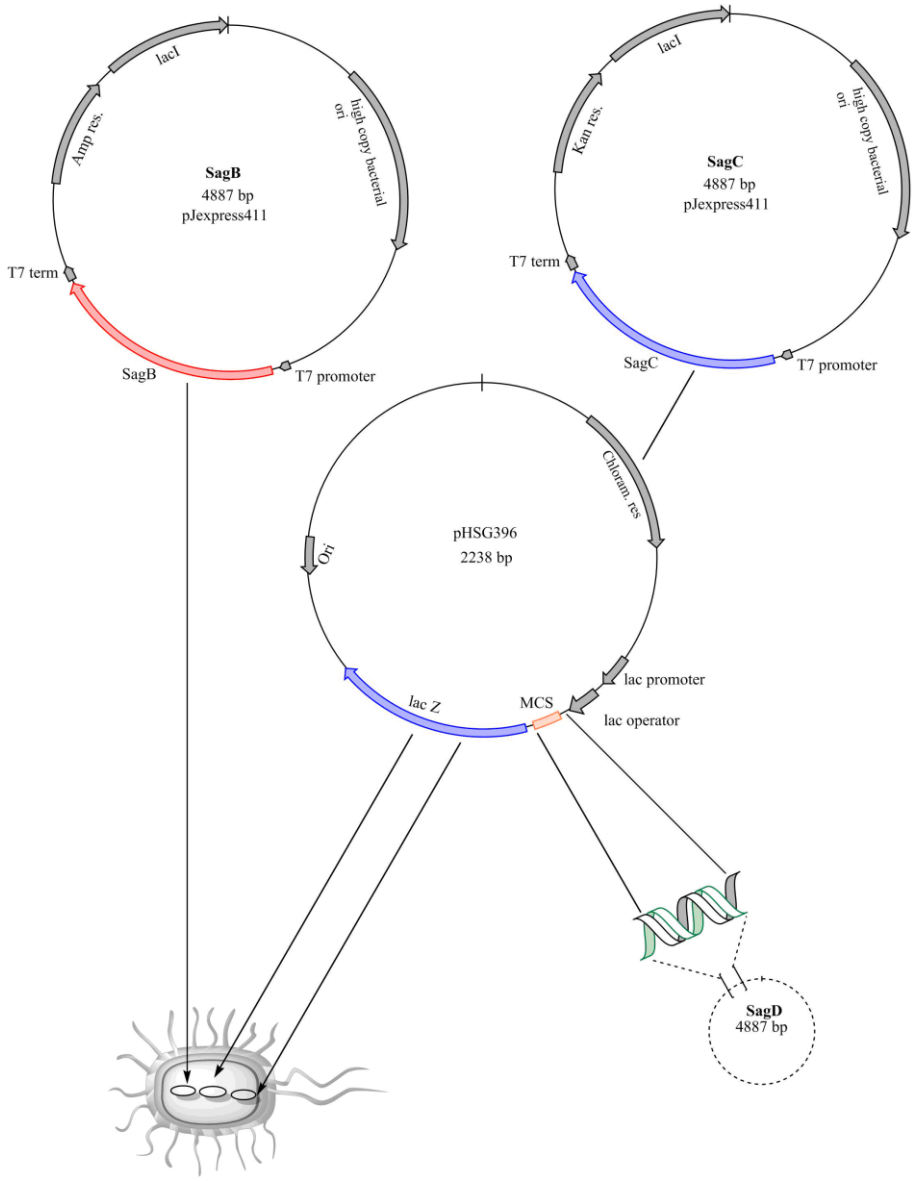


Figure 5.16: Schematic of SagBCD molecular cloning strategy 2. This diagram describes Strategy 2 for SagBCD co-expression in *E. coli*. This strategy involves the insertion of three individual plasmids (expressing specifically for SagB, SagC, and SagD) into BL21 (DE3) *E. coli* for co-expression experiments.

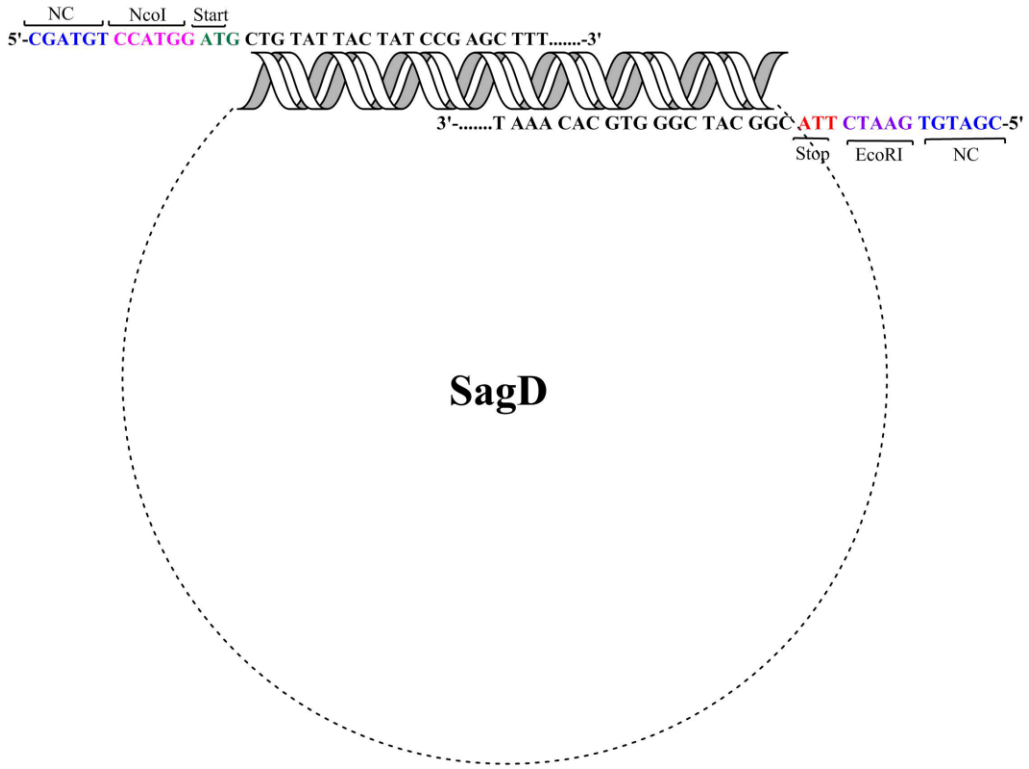


Figure 5.17: Schematic of primer design used for restriction site modification of SagD (Strategy 1). This diagram displays the strategy for introducing new restriction sites into the SagD plasmid prior to insertion of SagD reading frame into pETDuet™. This is done to ensure restriction site compatibility during ligation experiments.

5.5 REFERENCES

1. Molloy, Evelyn M., Paul D. Cotter, Colin Hill, Douglas A. Mitchell, and R. Paul Ross. "Streptolysin S-like Virulence Factors: The Continuing SagA." *Nature Reviews Microbiology* 9.9 (2011): 670-81. Print.
2. Lee, S. W., D. A. Mitchell, A. L. Markley, M. E. Hensler, D. Gonzalez, A. Wohlrab, P. C. Dorrestein, V. Nizet, and J. E. Dixon. "Discovery of a Widely Distributed Toxin Biosynthetic Gene Cluster." *Proceedings of the National Academy of Sciences* 105.15 (2008): 5879-884. Web
3. Mitchell, D. A., S. W. Lee, M. A. Pence, A. L. Markley, J. D. Limm, V. Nizet, and J. E. Dixon. "Structural and Functional Dissection of the Heterocyclic Peptide Cytotoxin Streptolysin S." *Journal of Biological Chemistry* 284.19 (2009): 13004-3012. Web.
4. Dunbar, Kyle L., Joel O. Melby, and Douglas A. Mitchell. "YcaO Domains Use ATP to Activate Amide Backbones during Peptide Cyclodehydrations." *Nature Chemical Biology Nat Chem Biol* 8.6 (2012): 569-75. Web.
5. Slabinski, Lukasz, Lukasz Jaroszewski, Ana P.C. Rodrigues, Leszek Rychlewski, Ian A. Wilson, Scott A. Lesley, and Adam Godzik. "The Challenge of Protein Structure Determination-lessons from Structural Genomics." *Protein Science* 16.11 (2007): 2472-482. Print.
6. Markley, Andrew, Emily Jensen, and Shaun Lee. "An Escherichia Coli-based Bioengineering Strategy to Study Streptolysin S Biosynthesis." *Analytical Biochemistry* 420 (2011): 191-93. Print.
7. Altschul, S., T. L. Madden, A. A. Schaffer, J. Zhang, W. Miller, and D. J. Lipman. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25.17 (1997): 3389-402. Web.
8. Fox, Jeffrey D., and David S. Waugh. "Maltose-Binding Protein as a Solubility Enhancer." *E. Coli Gene Expression Protocols* 205 (2003): 99-118. Print.
9. Dong, Aiping, Xiaohui Xu, Aled M. Edwards, Changsoo Chang, Maksymilian Chruszcz, Marianne Cuff, Marcin Cymborowski, Rosa Di Leo, Olga Egorova, Elena Evdokimova, Ekaterina Filippova, Jun Gu, Jennifer Guthrie, Alexandr Ignatchenko, Andrzej Joachimiak, Natalie Klostermann, Youngchang Kim, Yuri Korniyenko, Wladek Minor, Qiuni Que, Alexei Savchenko, Tatiana Skarina, Kemin Tan, Alexander Yakunin, Adelinda Yee, Veronica Yim, Rongguang Zhang, Hong Zheng, Masato Akutsu, Cheryl Arrowsmith, George V. Avvakumov, Alexey Bochkarev, Lars-Göran Dahlgren, Sirano Dhe-Paganon, Slav Dimov, Ludmila Dombrovski, Patrick Finerty, Susanne Flodin, Alex Flores, Susanne Gräslund, Martin Hammerström, Maria Dolores Herman, Bum-Soo Hong, Raymond Hui, Ida Johansson, Yongson Liu, Martina Nilsson, Lyudmila Nedyalkova, Pär Nordlund, Tomas Nyman, Jinrong Min, Hui Ouyang, Hee-won Park, Chao Qi, Wael Rabeh, Limin Shen, Yang Shen, Deepthi Sukumard, Wolfram Tempel, Yufeng Tong, Lionel Tresagues, Masoud Vedadi, John R. Walker, Johan Weigelt, Martin Welin, Hong Wu, Ting Xiao, Hong Zeng, and Haizhong Zhu. "In Situ Proteolysis for Protein Crystallization and Structure Determination." *Nature Methods* 4.12 (2007): 1019-021. Print.

10. Fersht, A. "The Folding of an Enzyme I. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding." *Journal of Molecular Biology* 224.3 (1992): 771-82. Print.
11. Walter, Thomas S., Christoph Meier, Rene Assenberg, Kin-Fai Au, Jingshan Ren, Anil Verma, Joanne E. Nettleship, Raymond J. Owens, David I. Stuart, and Jonathan M. Grimes. "Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization." *Structure* 14.11 (2006): 1617-622. Print.
12. Lund, A. H., M. Duch, and F. Skou Pedersen. "Increased Cloning Efficiency by Temperature-Cycle Ligation." *Nucleic Acids Research* 24.4 (1996): 800-01. Print.
13. Lilie, Hauke, Elisabeth Schwarz, and Rainer Rudolph. "Advances in Refolding of Proteins Produced in E. Coli." *Current Opinion in Biotechnology* 9.5 (1998): 497-501. Web.
14. Thermo, Scientific. *Pierce® Protein Refolding Kit*. Protein Refolding Protocol. N.p.: n.p., n.d. *Pierce® Protein Refolding Kit*. Thermo Scientific, 9 Feb. 2014. Web. 16 Mar. 2016.
15. Tolia, Niraj H., and Leemor Joshua-Tor. "Strategies for Protein Coexpression in Escherichia Coli." *Nature Methods Nat Meth* 3.1 (2006): 55-64. Web.
16. Die, I. M. Van, H. E. N. Bergmans, and W. P. M. Hoekstra. "Transformation In Escherichia Coli: Studies On The Role Of The Heat Shock In Induction Of Competence." *Microbiology* 129.3 (1983): 663-70. Web.
17. "Double Digest Finder." *Double Digest Finder*. New England Biolabs, n.d. Web. 18 Jan. 2016.
18. Sambrook, Joseph, E. F. Fritsch, and Tom Maniatis. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1989. Print.

APPENDIX A – DSF LIGAND SCREENING CONDITIONS

Table A.1. Hampton screen additives (DSF)
(*All solutions also contain 0.1 M Tris (pH 7.5)*)

A1	0.1M Barium Chloride
A2	0.1M Cadmium Chloride
A3	0.1M Calcium Chloride
A4	0.1M Cobalt(II) Chloride
A5	0.1M Copper(II) Chloride
A6	0.1M Magnesium Chloride
A7	0.1M Manganese(II) Chloride
A8	0.1M Strontium Chloride
A9	0.1M Yttrium(III) Chloride
A10	0.1M Zinc Chloride
A11	0.1M Iron(III) Chloride
A12	0.1M Nickel(II) Chloride
B1	0.1M Chromium(III) Chloride
B2	0.1M Praseodymium(III) Acetate
B3	1M Ammonium Sulfate
B4	1M Potassium Chloride
B5	1M Lithium Chloride
B6	2M Sodium Chloride
B7	0.5M Sodium Fluoride
B8	1M Sodium Iodide
B9	2M Sodium Thiocyanate
B10	1M Potassium Sodium Tartrate
B11	1M Sodium Citrate
B12	1M Cesium Chloride
C1	1M Sodium Malonate
C2	0.1M L-Proline
C3	0.1M Phenol
C4	30% DMSO
C5	0.1M Sodium Bromide
C6	30% 6-aminohexanoic acid
C7	30% 1,5-diaminopentane
C8	30% 1,6-diaminohexane
C9	30% 1,8-diaminooctane
C10	1M Glycine
C11	0.3M Glycyl-glycyl-glycine
C12	0.1M Taurine
D1	0.1M Betaine HCl
D2	0.1M Spermidine
D3	0.1M Spermidine x 4HCl
D4	0.1M Hexammine Cobalt(III) Chloride
D5	0.1M Sarcosine
D6	0.1M Trimethylamine HCl
D7	1M Guanidine HCl
D8	0.1M Urea
D9	0.1M β-Nicotinamide Adenine
D10	0.1M ATP
D11	0.1M TCEP HCl
D12	0.01M GSH
E1	0.1M EDTA
E2	5% Polyvinylpyrrolidone
E3	30% Dextran Sulfate Sodium Chloride
E4	40% Pentaerythritol ethoxylate
E5	10% PEG 3350
E6	30% D-(+)-Glucose
E7	30% Sucrose
E8	30% Xylitol
E9	30% D-Sorbitol
E10	12% Myo-inositol
E11	30% D-(+)-Trehalose
E12	30% D-(+)-Galactose
F1	30% Ethylene Glycol
F2	30% Glycerol
F3	3M NDSB-195
F4	2M NDSB-201
F5	2M NDSB-211
F6	2M NDSB-221
F7	1M NDSB-256
F8	0.15mM CYMAL-7
F9	20% Benzamidine
F10	5% LDAO, DDAO
F11	5% n-octyl-β-D-glucoside
F12	5% n-dodecyl-β-D-maltoside
G1	30% Trimethylmine-N-oxide
G2	30% 1,6-hexanediol
G3	30% 2-methyl-2,4-pentanediol
G4	50% PEG 400
G5	40% Jeffamine M-600
G6	40% 2,5-hexanediol
G7	40% 1,3-butanediol
G8	40% PEG 400
G9	30% 1,4-dioxane
G10	30% Ethanol
G11	30% 2-Propanol
G12	30% Methanol
H1	40% 1,4-butanediol
H2	40% tert-butanol
H3	40% 1,3-propanediol
H4	40% Acetonitrile
H5	40% Formamide
H6	40% 1-propanol
H7	5% Ethylacetate
H8	40% Acetone
H9	0.25% Dichloromethane
H10	7% 1-butanol
H11	40% 2,2,2-trifluoroethanol
H12	40% 1,1,1,3,3,3-hexafluoro-2-propano

Table A.2: DSF ligand screen 2
(*All solutions also contain 0.05 M Tris (pH 7.5)*)

A1	0.1M Valine
A2	0.1M Alanine
A3	0.1M Arginine
A4	0.1M Malate
A5	0.1M Sodium Citrate
A6	0.1M Threonine
A7	0.1M Glutamine
A8	0.1M Isoleucine
A9	0.1M Succinate
A10	0.1M Sodium Pyruvate
A11	0.1M Glucose
A12	0.1M Lysine
B1	0.1M Asparagine
B2	0.1M Homoserine
B3	0.1M Cytosine
B4	0.1M Glutamate
B5	0.1M Proline
B6	0.1M Taurine
B7	0.1M Nicotinamide
B8	0.1M Guanine
B9	0.1M Glycine
B10	0.1M Serine
B11	0.1M Methionine
B12	0.1M 4-methylcatechol
C1	0.1M Citrate
C2	0.1M Nicotinamide
C3	0.01M Magnesium Chloride
C4	0.1M Leucine
C5	0.01M Zinc Chloride
C6	0.01M Iron(III) Chloride
C7	0.1M Sodium Pyrophosphate
C8	0.1M Thymine
C9	0.1M Sucrose
C10	0.01M Calcium Chloride
C11	0.1M Spermidine
C12	0.1M 2,6-Pyridine dicarboxylic acid methyl ester
D1	0.1M Tryptophan
D2	0.1M Sodium Cholate
D3	0.1M Sodium Pyrophosphate
D4	0.01M Manganese(II) Chloride
D5	0.1M 2,6-Pyridine dicarboxylic acid ethyl ester
D6	0.1M Aspartate
D7	0.1M Adenine
D8	0.1M Xylitol
D9	0.01M Cobalt(II) Chloride
D10	0.1M 2,6-Pyridine dicarboxylic acid
D11	0.1M Phenylalanine
D12	0.1M Uracil
E1	0.1M Mercaptosuccinate
E2	0.1M γ-amino-butyric acid
E3	0.1M 8-hydroxy-2-quinole carboxylic acid
E4	0.1M Pimelic acid
E5	0.1M Sodium Benzene 1,3-disulfonate
E6	0.1M 6-oxo-pipecolinic acid
E7	0.1M Kanamycin
E8	0.1M 5-amino-isophthalic acid
E9	0.1M Chelidanic acid
E10	0.1M 2-pyridine sulfonic acid
E11	0.1M 2-proline
E12	0.1M Gly-gly
F1	0.1M 3,3',5,5'-tetracarboxyl diphenalmethane
F2	0.1M 5-nitrofuroic acid
F3	0.1M 4,5-dibromo-2-furoic acid
F4	0.1M 5-sulfoisophthalic acid
F5	0.1M Betaine
F6	0.1M isocitric acid lactone
F7	0.1M 3-aminobenzene sulfonic acid
F8	0.1M Chelidonic acid
F9	0.1M 1,3,5-benzene tricarboxylic acid
F10	0.1M isocitric acid
F11	0.1M p-coumaric acid
F12	0.1M 2,5-thiophene dicarboxylic acid
G1	0.1M Phosphoserine
G2	0.1M Tartaric acid
G3	0.1M Arabinose
G4	0.1M Chloramphenicol
G5	0.1M Naringin
G6	0.1M Caffeic acid
G7	0.1M 2,4-pyridine dicarboxylic acid
G8	0.1M 2,3-pyridine dicarboxylic acid
G9	0.1M 3,5-pyridine dicarboxylic acid
G10	0.1M 2,5-pyridine dicarboxylic acid
G11	0.1M Tyrosine
G12	0.1M Rutin
H1	0.1M Morin
H2	0.01M Copper(II) Chloride
H3	0.1M Maltose
H4	0.1M AMP
H5	0.1M Myo-inositol
H6	0.1M Ascorbic acid
H7	0.1M Thiamine
H8	0.1M Pimelic acid
H9	0.1M 3-aminobenzene sulfonic acid
H10	0.1M Melamine
H11	0.1M 2,5 FDC
H12	0.1M Cystine

Table A.3: DSF ligand screen 3
(*All solutions also contain 0.1 M Tris (pH 7.5)*)

A1	30% Methanol
A2	30% Ethanol
A3	30% DMSO
A4	40% Acetone
A5	30% Glycerol
A6	30% Ethylene Glycol
A7	0.1M Bromosuccinic acid
A8	0.1M Cytidine
A9	0.1M Cytidine in 100% Ethanol
A10	0.1M Sodium Deoxycholate
A11	0.1M Sodium Chloride
A12	0.1M Trimesic acid
B1	0.1M Histidine HCl
B2	0.1M 1,4-butanediol
B3	0.1M Neopentyl Alcohol
B4	0.1M Cadmium Chloride
B5	0.1M Malonic acid
B6	0.1M 3,4-dihydroxybenzoic acid
B7	0.1M CHES
B8	0.1M MOPS
B9	0.1M L-norvaline
B10	0.1M 1,6-hexanediol
B11	0.1M Lanthanum Chloride Heptahydrate
B12	0.1M Vanillin
C1	0.1M Kanamycin Sulfate
C2	0.1M N,N-dimethylglycine
C3	0.1M Salicylic acid
C4	0.1M 3-hydroxy-2-naphthoic acid
C5	0.1M Tris(hydroxymethyl)aminomethane
C6	0.1M CAPS
C7	0.1M Ethylene Glycol
C8	0.1M Chloroquine
C9	0.1M Potassium Thiocyanate
C10	0.1M Fast Garnet GBC Salt
C11	0.1M N,N-dimethyldecylamine-N-oxide
C12	0.1M 1,4-butanediphosphonic acid
D1	0.1M Lithium Acetate
D2	0.1M Magnesium Nitrate
D3	0.1M Fast Violet B Salt
D4	0.1M β-NAD
D5	0.1M α-lactose monohydrate
D6	0.1M Guanidine Thiocyanate
D7	0.1M Sodium Bisulfate
D8	0.1M BOC-L-serine
D9	0.1M Glutathione (reduced)
D10	0.1M PIPES
D11	0.1M Sodium Bicarbonate
D12	0.1M Thiourea
E1	0.1M α-naphthyl acid phosphate
E2	0.1M Streptomycin Sulfate
E3	0.1M Hexamine Cobalt(III) Chloride
E4	0.1M Magnesium Sulfate
E5	0.1M Lithium Sulfate
E6	0.1M 4-dimethylaminopyridine
E7	0.1M S-(-)-2-bromopropionic acid
E8	0.1M O-phospho-DL-serine
E9	0.1M Nickelous Acetate
E10	0.1M Tetraamine Copper(II) Sulfate
E11	0.1M Potassium Tetrachloroplatinate
E12	0.1M Ammonium Cerium(IV) Nitrate
F1	0.1M Chromium Acetate
F2	0.1M Trifluoroacetic acid Silver Salt
F3	0.1M Nicotinic acid
F4	0.1M Trimethyl Lead Acetate
F5	0.1M Propylenediphosphonic acid
F6	0.1M Nickel Sulfate
F7	0.1M Tetraethylpropylene-1,3-diphosphonic acid
F8	0.1M 1,6-hexane bis-phosphonic acid
F9	0.1M Cesium Fluoride
F10	0.1M Strontium Nitrate
F11	0.1M Cobaltous Sulfate
F12	0.1M Glutathione
G1	0.1M Cerous Chloride
G2	0.1M Sodium Molybdate(III) dehydrate
G3	0.1M Thallium(I) Nitrate
G4	0.1M Cesium Chloride
G5	0.1M Iodoacetamide
G6	0.1M Phosphonoacetic acid
G7	0.1M 2,6-naphthalenedisulfonic acid disodium salt
G8	0.1M Propionic acid
G9	0.1M Calcium Acetate
G10	0.1M 1,1,3,3-tetramethylurea
G11	0.1M Adonitol
G12	0.1M Nicotinamide
H1	0.1M GMP
H2	0.1M Cacodylic acid
H3	0.1M Calcium Formate
H4	0.1M Lithium Citrate
H5	0.1M Imidazole
H6	0.1M Lithium Chloride
H7	30% Dextrose
H8	0.1M Sodium Sulfate
H9	0.1M Thymidine
H10	0.1M Potassium Hydrogen Phthalate
H11	0.1M Ferrous Sulfate
H12	0.1M Hydroxylaminohydrochloride

Table A.4: DSF ligand screen 4
(*All solutions also contain 0.1 M Tris (pH 7.5)*)

A1	0.1M ADA (N-2-acetamide-2-iminodiacetic acid)
A2	0.1M trans-4,5-dihydroxy-1,2-dithiane
A3	0.1M 1-naphthalene acetic acid
A4	0.1M β-alanine
A5	0.1M Potassium Chloride
A6	0.1M Cerium(III) Chloride Heptahydrate
A7	0.1M Lithium Fluoride
A8	0.1M 5,5'-dithiobis(2-nitrobenzoic acid)
A9	0.1M 1,2,4,5-benzenetetracarboxylic acid
A10	0.1M Cyclohexylsuccinic acid
A11	0.1M 6-(methoxycarbonyl) picolinic acid
A12	0.1M CHAPS
B1	0.1M trans-cinnamic acid
B2	0.1M Sulfonic acid
B3	0.1M Sodium 3-sulfobenzoate
B4	0.1M 8-anilino-1-naphthalene sulfonic acid
B5	0.1M N-bromosuccinimide
B6	0.1M Terephthalic acid
B7	0.1M Chlorosuccinic acid
B8	0.1M Guanidine HCl
B9	2% Triton-X114
B10	0.1M L-serine
B11	0.1M Dextrose Anhydrous
B12	0.1M Sodium Tartrate Dihydrate
C1	0.1M Sodium Acetate Trihydrate
C2	0.1M PEG 400
C3	10% Poly(vinyl Alcohol)
C4	0.1M 1,2-dihydroxybenzene
C5	0.1M Sodium Malonate Dibasic Monohydrate
C6	0.1M 5-aminovaleric acid
C7	0.05M PEG 8000
C8	0.1M PEG 200
C9	0.1M NDSB-221
C10	0.1M NDSB-201
C11	0.1M 2,2-dimethylsuccinic acid
C12	0.05M PEG 6000
D1	0.1M PEG 4000
D2	0.1M NDSB-256
D3	0.1M Glycerol
D4	0.1M MES Free Acid Monohydrate
D5	0.1M PEG 2000 MME
D6	0.1M PEG 5000 MME
D7	0.1M NDSB-195
D8	0.1M Phthalic acid
D9	0.1M 2-methyl-2,4-pentanediol
D10	0.1M PEG 3350
D11	0.1M PEG 1000
D12	0.1M 2-hydroxy-2-(2-isopropoxy-2-oxoethyl) succinic acid
E1	0.1M 6-(ethoxylcarbonyl) Picolinic acid
E2	0.1M EDTA
E3	0.1M Potassium Sulfate
E4	0.1M Sodium Carbonate
E5	0.06M N-bromosuccinic acid in 40% Acetone
E6	0.07M Thymol in 30% Ethanol
E7	0.07M 2-adamantanone in 30% Methanol
E8	0.07M N-bromosuccinic acid in 30% DMSO
E9	0.01M 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)
E10	0.07M Sodium Tetraborate in 30% Ethylene Glycol
E11	0.1M Mellitic acid
E12	0.1M methyl-6-hydroxymethyl-2-carboxylate pyridine
F1	0.07M Decanoic acid in 30% Ethanol
F2	0.01M Tetrasodium Pyrophosphate
F3	0.07M Thymine in 30% Methanol
F4	0.01M Calcium Gluconate
F5	0.01M D-2-phosphoglyceric acid
F6	0.07M Hexadecyltrimethylammonium bromide in 30% Ethanol
F7	0.07M 1,3-dicyclohexyl-carbodiimide in 30% Ethanol
F8	0.01M Asparagine
F9	0.01M Calcium Sulfate
F10	0.07M 2,4-dibromophenol in 30% Ethanol
F11	0.01M Adenine
F12	0.07M 3-amino-4-hydroxybenzoic acid in 30% Ethanol
G1	0.06M 3-amino-4-hydroxybenzoic acid in 40% Acetone
G2	0.01M Phenylurea
G3	0.06M Chloramphenicol in 40% Acetone
G4	0.07M Thymine in 30% DMSO
G5	0.06M 3-indoleacetic acid in 40% Acetone
G6	0.07M Melamine in 30% Glycerol
G7	0.07M 5-iodosalicylic acid in 30% Ethanol
G8	0.07M 4-aminobiphenyl in 30% DMSO
G9	0.01M Magnesium Citrate
G10	0.07M Caffeine in 30% Ethanol
G11	0.07M Quinine HCl in 30% Ethanol
G12	0.07M Rhodanine in 30% Ethanol
H1	0.1M (2R,6S)-2,6-piperdinedicarboxylate
H2	0.1M 8-aminooctanoic acid
H3	0.1M O-phospho-L-serine
H4	0.1M Diphenylamino-4-sulfonic acid
H5	0.1M Phytic acid
H6	0.1M TCEP HCl
H7	0.01M Samarium Sulfate
H8	0.01M Dibutylthiourea
H9	0.07M 2,6-dibromophenol in 30% Ethanol
H10	0.07M Kinetin in 30% DMSO
H11	0.07M 1-phenyl-2-thiourea in 30% Ethanol
H12	0.01M Thallium Chloride